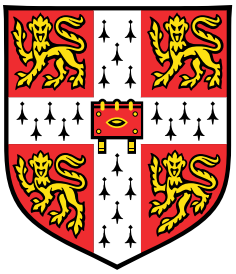
Measurement of electron-electron interactions in one dimension with tunnelling spectroscopy



Yiqing Jin

Semiconductor Physics Group, Cavendish Laboratory
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Clare Hall

March 2020

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Yiqing Jin
March 2020

Abstract

Measurement of electron-electron interactions in one dimension with tunnelling spectroscopy

Yiqing Jin

The properties of fermions in one dimension are markedly different from those in higher dimensions. In two or three dimensions, the many-body behaviour of fermions is described by the Fermi liquid model, whose most prominent result is that, instead of studying the behaviour of individual particles, it is better to describe the emergent phenomena of the whole system, known as quasiparticles. These are collective modes of excitations born out of particle-particle interactions, which resemble the constituent single particles, albeit with renormalised mass. In one dimension, however, the formation of such a correlated state is unfavourable and the quasiparticles are completely absent. Instead, 1D fermions are described by a model known as the Tomonaga-Luttinger liquid. Central to this model is the process of bosonisation, a mathematical technique which maps the excitation of fermions to bosonic modes of spin-density and charge-density waves. As a prerequisite of the bosonisation technique, the fermionic single-particle dispersion is approximated as a linear function. Consequently, the Luttinger liquid model is limited in its scope of applicability. Specifically, it is valid only in the vicinity of the Fermi points, where the excitation energy is low.

Recent advancement in the theoretical field has expanded our understanding of 1D interacting fermions beyond the limit of linearity. The mobile-impurity model, proposed by L. Glazman and his co-authors, is a generalised model of 1D fermions which extends the scope of applicability of the Tomanaga-Luttinger model, accounting for the curvature of single-particle dispersion relations. It extends the region of validity of the tradition TLL theory from the neighbourhood of the Fermi points to a band which follows the single-particle dispersion. This thesis presents the results of an experimental work which provides direct evidence of non-linear behaviour predicted by the new model. It revolves around the mea-

surement of the spectral function of electrons in surface-gate-defined quantum wires on a double-quantum-well, GaAs/AlGaAs device, and the comparison and fitting of the result to the non-linear theory. The spectral function is a theoretical concept frequently used in discussions on many-body problems. Its physical interpretation makes it closely related to experimental observables, such as the tunnelling current between the double quantum wells in this work.

This thesis consists of three parts. Part one concerns the theories of the various many-body models and the experimental technique: Chapter 1 briefly introduces the concept of Fermi liquid, before focusing on a detailed account of the Luttinger-liquid model. The mobile-impurity model will be introduced in the framework of the Luttinger liquid model afterwards, followed by a review of experimental works reported in the literature on the study of TLL physics. In Chapter 2, the principles of the experimental tunnelling spectroscopy technique are described. The second part of the thesis details the structure of the test device used in the experiment and its fabrication. In particular, we will discuss the intricacies of the three-dimensional EBL fabrication process developed specifically for this project. The setup and running of the measurement circuit will also be discussed in this part of the thesis, in the context of a prerequisite sample characterisation measurement. In the final part of the work, data from the main experiment will be presented and analysed, in search for the electron-electron interaction effects which are predicted by the theoretical discussions in the first chapter.

Acknowledgements

I would like to acknowledge my profound gratitude to my supervisor Prof. Chris Ford for his unfading patience and tireless guidance throughout my time at the SP group. His sharp intellect and immense academic expertise have always been, and remain to be my enduring inspiration. I am especially obliged to thank Dr. Oleksandr Tsyplyatyev at the Goethe-Universität Frankfurt for his indispensable contribution to this project, and for his unflagging support and assistance. Thanks to the innumerous hours of discussions with him, many intricate concepts in the many-body theory were elucidated to a non-theorist like me. I must also thank Jonathan Griffiths at the Cavendish Laboratory for his critical help in developing the three-dimensional EBL technique which played a pivotal role in the project.

I am deeply indebted to the assistance of my colleges at Cavendish over the course of the project. I am grateful to Dr. Maria Moreno for her help in the setup and running of a major part of the experiments, and to Dr. Anne Anthore for her help in analysis and troubleshooting the scripts for numerical calculations. Their keen insights have contributed greatly towards our understanding of the experimental data. In addition, this work would have been impossible without the help of Dr. Ian Farrer, whose expertise in MBE growth was instrumental in securing for us the necessary wafer material for extensive device fabrication.

I would also like to thank my many fellow research students at SP for their generous support. In particular, I wish to thank Wooi Kiat Tan for his unreserved and selfless assistance throughout the years, as well as for his cherished friendship. I would also like to thank Ankita Anirban for her assistance in the gathering and analysis of data in the later parts of the project.

Finally, I must thank my parents and wife Shengyang for their unconditional love and unwavering support throughout the long years of my studies. Without their motivation this work would never have been possible.


To my loving parents & Shengyang

"Even his griefs are a joy long after to one that remembers all that he wrought and endured."

— Homer, *The Odyssey*

Table of contents

Chapter 1

The one-dimensional many-body problem

1.1 The Fermi liquid

Fermi-liquid theory, proposed by Lev Landau in 1956, is a phenomenological theory which provides an approximate solution to many-body fermionic systems at low temperatures. Originally developed to explain the Fermi-gas-like properties of liquid ^{3}He, the theory is applicable to other fermionic systems such as conducting electrons in metals and semiconductors. It is now regarded as a paradigm in the study of the many-body problem.

Despite being a liquid (whose constituent particles should be in constant contact) at very low temperature, liquid ^{3}He's heat capacity resembles that of an ideal gas (whose particles do not interact). Landau considered an explanation to this paradoxical observation. In a free gas, any excited state can be specified by the deviation of the momentum distribution function from the ground state

$$\delta n_{\boldsymbol{k},\sigma} = n_{\boldsymbol{k},\sigma} - n^0_{\boldsymbol{k},\sigma}. \tag{1.1}$$

The distribution function, $n_{\boldsymbol{k},\sigma}$, specifies whether a quantum state of wave number $\boldsymbol{k}$ and spin σ is occupied. For fermions in the ground state, $n^0_{\boldsymbol{k},\sigma}$ is a step function which equals to 1 for $k \leq k_\mathrm{F}$ and 0 for $k > k_\mathrm{F}$. The change in total energy caused by the excitation is given by a functional of the deviation

$$\delta E[\delta n] = E[n] - E_0 = \sum_{\boldsymbol{k},\sigma} \epsilon_{\boldsymbol{k}}\, \delta n_{\boldsymbol{k},\sigma}\,, \tag{1.2}$$

where $\epsilon_{\boldsymbol{k}} = \hbar k^2/2m$ is the energy of a free particle. From this expression, one can define the energy of a particle as the functional derivative

$$\epsilon_{\boldsymbol{k}} = \frac{\delta E[n]}{\delta n_{\boldsymbol{k},\sigma}}. \tag{1.3}$$

An elementary excitation is formed by adding or removing a particle from the ground state. An arbitrary excited state may consist of a number of such particle-hole pairs. Close to the Fermi surface, one may expand the dispersion relation as

$$\epsilon_{\boldsymbol{k}} = \epsilon_{\mathrm{F}} + v_{\mathrm{F}}\hbar(k - k_{\mathrm{F}}) + \mathcal{O}\big((k - k_{\mathrm{F}})^2\big), \tag{1.4}$$

where k_{F} and ϵ_{F} are the Fermi wavevector and energy, respectively, and $v_{\mathrm{F}} = \nabla_{\boldsymbol{k}}\epsilon_{\boldsymbol{k}}|_{\boldsymbol{k}=\boldsymbol{k}_{\mathrm{F}}} = \hbar\, k_{\mathrm{F}}/m$ is the Fermi velocity.

The key idea of Landau's theory of Fermi liquid lies upon the concept of an adiabatic process: consider a hypothetical scenario in which particle-particle interaction of a Fermi gas is slowly turned on from zero, and the gas gradually becomes liquid. During this process, the eigenstate of the system evolves continually from that of a free gas into that of a liquid. Crucially, there always exists a one-to-one correspondence between the initial non-interacting state and the final interacting state. According to the Luttinger theorem [97], the volume of the Fermi surface is not changed by interaction. Consequently, the Fermi momentum remains the same after the interaction is turned on. An excited state of the interacting system can be generated by adding a particle with $k > k_{\mathrm{F}}$ to the non-interacting ground state (the Fermi sphere) and adiabatically turning on the interaction. Such a state is known as a quasiparticle (an excited state formed by adding a hole with $k < k_{\mathrm{F}}$ is called a quasihole). Since momentum and spin are conserved throughout the adiabatic process, each quasiparticle can be labelled by the same quantum numbers as the non-interacting particle.

Interaction causes scattering and results in a finite lifetime τ of quasiparticles. For the quasiparticle model to be valid, the corresponding relaxation rate $\hbar/\tau$ must be smaller than the energy of the quasiparticle ϵ. Consider a particle above the Fermi surface with momentum $\boldsymbol{p}_1$: after it interacts with another particle below (and closer to) the Fermi surface with momentum $\boldsymbol{p}_2$, two new particles will appear above the Fermi surface with momentum $\boldsymbol{p}_3$ and $\boldsymbol{p}_4$ (Figure 1.1). This process can be interpreted as the particle with $\boldsymbol{p}_1$ scattering and decaying into particles of momenta $\boldsymbol{p}_3$ and $\boldsymbol{p}_4$, plus a hole of $\boldsymbol{p}_2$. According to Fermi's golden rule, the scattering rate from an initial state i to a final state f is

$$\frac{1}{\tau} = \frac{2\pi}{\hbar}\sum_{\mathrm{f}} \left|T_{i\to f}\right|^2 \rho(E_{\mathrm{f}}) \propto \int \mathrm{d}\boldsymbol{p}_2\, \mathrm{d}\boldsymbol{p}_3\, \delta(\epsilon_1 + \epsilon_2 - \epsilon_3 - \epsilon_4), \tag{1.5}$$

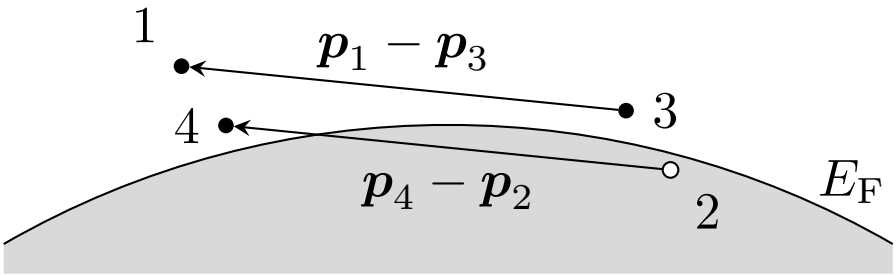


FIGURE 1.1 A particle close to the Fermi surface (label 1) scatters into a hole (label 2), creating two new particles (labels 3 and 4). Conservation of energy and momentum limits the momenta of the new particles by demanding the two arrows to be identical. If $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are close to the Fermi surface, $\boldsymbol{p}_3$ and $\boldsymbol{p}_4$ must also be close.

where in the second part of the relation we have used the delta function as the density of the final states, because electrons and quasiparticles obey Pauli's exclusion principle. We assume that the scattering matrix element, $\left|T_{i\to f}\right|$, is constant. The particle energies are given by $\epsilon_i = p_i^2/2m - E_\mathrm{F}$ and the momenta are bounded by the conditions $p_2 < E_\mathrm{F}$, $p_3 > E_\mathrm{F}$, and $p_4 = \left|\boldsymbol{p}_1 + \boldsymbol{p}_2 - \boldsymbol{p}_3\right| > E_\mathrm{F}$. Given $\epsilon_1 + \epsilon_2 = \epsilon_3 + \epsilon_4 > 0$, if ϵ_1 is small (near the Fermi surface), then the negative ϵ_2 is also small. This means $\boldsymbol{p}_2$ is limited within a narrow shell around the Fermi surface, and only a fraction $\sim \epsilon_1/E_\mathrm{F}$ of states can participate in the scattering. Additionally, since the condition $\boldsymbol{p}_1 - \boldsymbol{p}_3 = \boldsymbol{p}_2 - \boldsymbol{p}_4$ must always be satisfied, it means the two vectors in Figure 1.1 must always be identical. Therefore $\boldsymbol{p}_3$, $\boldsymbol{p}_4$ are also limited to inside a narrow shell close to the Fermi surface. As a result, the integral in equation 1.5 is on the order of $(\epsilon_1/E_\mathrm{F})^2$. This means $\tau \propto (\epsilon - E_\mathrm{F})^{-2}$, i. e. a quasiparticle is more stable the closer it is to the Fermi surface, because of reduced scattering. For this reason, we see that the Fermi liquid theory is valid only in the low-energy regime. The suppression of scattering near the Fermi surface explains why the Fermi liquid's behaviour resembles a free gas despite the presence of interaction.

The dispersion of a quasiparticle can be written as

$$\epsilon_{\boldsymbol{k}} = \mu + v\hbar(k - k_\mathrm{F}) + \mathcal{O}\big((k - k_\mathrm{F})^2\big), \tag{1.6}$$

where $v = \hbar v_\mathrm{F}/m^*$ and μ is the chemical potential. Essentially the effect of the interaction is captured by the single parameter, m^*, which is a renormalised mass. A quasiparticle can be visualised as a dressed particle: a fermion surrounded by a cloud of disturbance which arises due to interaction. The (increased) effective mass reflects that the fermion drags the cloud with it as it moves through the Fermi sea.

In order to find the quasiparticle excitation energies, one needs to expand the energy functional to the second order

$$\delta E[\delta n] = \sum_{\boldsymbol{k},\sigma} \epsilon_{\boldsymbol{k}}\, \delta n_{\boldsymbol{k}}, \sigma + \frac{1}{2V} \sum_{\boldsymbol{k},\sigma,\boldsymbol{k}',\sigma'} f(\boldsymbol{k},\boldsymbol{k}')\, \delta n_{\boldsymbol{k},\sigma}\, \delta n_{\boldsymbol{k}',\sigma'}\,. \tag{1.7}$$

The intensive quantity, $f(\boldsymbol{k},\boldsymbol{k}')\big/V = \delta^2 E[n]\big/\delta n_k \delta n_{k'}$, defined in terms of the second order functional derivative of the total energy, is known as the Landau function. It is related to the interaction between quasiparticles and is central to the Fermi liquid theory. The energy of a quasiparticle can be defined as

$$\tilde{\epsilon}_{\boldsymbol{k}} = \frac{\delta E[\delta n]}{\delta n_{\boldsymbol{k},\sigma}} = \epsilon_{\boldsymbol{k}} + \frac{1}{V} \sum_{\boldsymbol{k}',\sigma'} f(\boldsymbol{k},\boldsymbol{k}')\, \delta n_{\boldsymbol{k}',\sigma'}\,. \tag{1.8}$$

This expression shows that the quasiparticle energy is a functional of δn only. The energy expression is the starting point from which many equilibrium properties of the Fermi liquid, such as the specific heat, compressibility and spin susceptibility, can be derived [21]. These quantities turn out to be of similar form to those of the free Fermi gas, with minor deviation such as a renormalised mass and additional terms with Landau-function dependencies.

1.2 Green's function formalism

1.2.1 Definitions

Landau's original solution to the Fermi liquid was obtained by expanding the macroscopic quantity of the total energy to second order. Later literature showed that a detailed, microscopic approach using the language of second quantisation can lead to equivalent results. This approach involves the use of the Green's function method. Originally devised as a tool for solving linear differential equations, the Green's function is a powerful formalism which can be generalised to the theory of linear operators. In the 1960s it was introduced into quantum field theory. We begin a review of the application of Green's function formalism to the many-body problem by introducing the correlation function, defined as the average of the product for any two operators

$$C(A(r_1), B(r_2)) = \langle A(r_1) B(r_2) \rangle = \langle \Psi | AB | \Psi \rangle\,, \tag{1.9}$$

where $|\Psi\rangle$ is the n-body state of the system in the Hilbert space. The Green's function is defined as the time-ordered correlation function between two operators

$$G_{AB}(t_1; t_2) = -i \langle T[A(t_1), B(t_2)] \rangle , \tag{1.10}$$

where T is the time-ordering operator, whose definition is

$$T[A(t_1), B(t_2)] = \Theta(t_1 - t_2)A(t_1)B(t_2) + \varepsilon\Theta(t_2 - t_1)B(t_2)A(t_1), \tag{1.11}$$

where $\Theta(t) = 1$ if $t > 0$, and $\Theta(t) = 0$ otherwise (i. e. $\Theta(t)$ is the Heaviside step function). Essentially, T sequences its operands in time to take place from right to left in the written order. The factor ε accounts for the exchange parity: it equals to 1 if A, B are bosonic operators, and –1 if they are fermionic.

An important Green's function is the one between single-particle field operators

$$G(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) = -i \left\langle T[c(\boldsymbol{x}_1, t_1), c^\dagger(\boldsymbol{x}_2, t_2)] \right\rangle , \tag{1.12}$$

where $c(\boldsymbol{x}_i, t_i)$ and $c^\dagger(\boldsymbol{x}_i, t_i)$ are operators which annihilate/create single particles at position $\boldsymbol{x}_i$ and time t_i. In the Heisenberg representation, operators evolve in time according to

$$A(t) = \mathrm{e}^{iHt/\hbar} A(0) \mathrm{e}^{-iHt/\hbar}, \tag{1.13}$$

where H is the Hamiltonian. Therefore, in the case of $t_1 > t_2$:

$$G(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) = -i \left\langle \mathrm{e}^{t_1 H} c(\boldsymbol{x}_1) \mathrm{e}^{-(t_1 - t_2)H} c^\dagger(\boldsymbol{x}_2) \mathrm{e}^{-t_2 H} \right\rangle . \tag{1.14}$$

We have simplified the above expression by making $\hbar = 1$, which will be assumed for the rest of the chapter unless stated otherwise. Equation 1.14 can be interpreted as the overlap (dot product) between two states: the 'ket' state is obtained by first evolving the system's ground state for t_2 in time, then adding a particle at position $\boldsymbol{x}_1$. The 'ket' state is then evolved further for a time interval of $t_1 - t_2$; the 'bra' state is obtained by evolving the system for t_1 in time, and removing a particle at $\boldsymbol{x}_2$. In essence, it is the probability amplitude of a system which has a particle added at $(\boldsymbol{x}_2, t_2)$ and another removed at $(\boldsymbol{x}_1, t_1)$. By combining Equation 1.11 and 1.12, we see that the Green's function can be separated into a 'retarded' and an 'advanced' part:

$$\begin{aligned} G^{\mathrm{R}}(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) &= -i \left\langle \Theta(t_1 - t_2) c(\boldsymbol{x}_1, t_1) c^\dagger(\boldsymbol{x}_2, t_2) \right\rangle ; \\ G^{\mathrm{A}}(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) &= -i\varepsilon \left\langle \Theta(t_2 - t_1) c(\boldsymbol{x}_1, t_1) c^\dagger(\boldsymbol{x}_2, t_2) \right\rangle . \end{aligned} \tag{1.15}$$

They are named thus because the averages are taken over the positive and negative half of the time domain, respectively.

Generally, if the Hamiltonian is time-independent and if the system is translationally invariant, the Green's function is dependent only on the temporal and spatial separation $t = t_1 - t_2$, $\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2$. Therefore,

$$G(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) = G(\boldsymbol{x}, t) = -i \left\langle T \left[c(\boldsymbol{x}, t), c^\dagger(0,0) \right] \right\rangle . \tag{1.16}$$

When such symmetries exist, one can use a Fourier transform to calculate Green's function between the momentum/frequency and position/time domains

$$\begin{aligned} G(\boldsymbol{k}, \omega) &= \int \mathrm{d}t \int \mathrm{d}^3\boldsymbol{x}\, \mathrm{e}^{-i(\boldsymbol{k}\boldsymbol{x} - \omega t)} G(\boldsymbol{x}, t) \\ G(\boldsymbol{x}, t) &= \int \frac{\mathrm{d}\omega}{2\pi} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3} \mathrm{e}^{i(\boldsymbol{k}\boldsymbol{x} - \omega t)} G(\boldsymbol{k}, \omega). \end{aligned} \tag{1.17}$$

The single-particle field operators and their counterparts in momentum space (e. g. $c^\dagger_{\boldsymbol{k}}$, which creates a particle of momentum $\boldsymbol{k}$) have the relations

$$c^\dagger(\boldsymbol{x}) = \frac{1}{\sqrt{V}} \sum_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\boldsymbol{x}} \qquad c(\boldsymbol{x}) = \frac{1}{\sqrt{V}} \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \mathrm{e}^{i\boldsymbol{k}\boldsymbol{x}}, \tag{1.18}$$

where V is the volume of the system. Substituting the above into Equation 1.17 yields

$$\begin{aligned} G(\boldsymbol{k}, \omega) &= -i \int \mathrm{d}t \int \mathrm{d}^3\boldsymbol{x}\, \mathrm{e}^{-i(\boldsymbol{k}\boldsymbol{x} - \omega t)} \left\langle T \left[c(\boldsymbol{x}, t) c^\dagger(0,0) \right] \right\rangle \\ &= -i \int \mathrm{d}t \int \mathrm{d}^3\boldsymbol{x}\, \mathrm{e}^{-i(\boldsymbol{k}\boldsymbol{x} - \omega t)} \left\langle T \left[\frac{1}{V} \sum_{\boldsymbol{k}'} \mathrm{e}^{i\boldsymbol{k}'\boldsymbol{x}} c_{\boldsymbol{k}'}(t) \sum_{\boldsymbol{k}''} c^\dagger_{\boldsymbol{k}''} \right] \right\rangle . \end{aligned} \tag{1.19}$$

According to the fermion anti-commutation relationship $\left\{ c_i, c_j^\dagger \right\} = \delta_{ij}$, the cross terms (i. e. when $i \neq j$) in the expansion of the double sum $\sum_{\boldsymbol{k}'} \sum_{\boldsymbol{k}''} c_{\boldsymbol{k}'}(t) c^\dagger_{\boldsymbol{k}''}$ vanish. It can therefore be simplified as the single sum $\sum_{\boldsymbol{k}'} c_{\boldsymbol{k}'}(t) c^\dagger_{\boldsymbol{k}'}$. Furthermore, we can replace $\sum_{\boldsymbol{k}}$ with $\frac{V}{(2\pi)^3} \int \mathrm{d}\boldsymbol{k}$ in the continual limit (where the prefactor ensures the number of states remains the same in a k-volume). Recall that the delta function has the integral representation:

$$\frac{1}{(2\pi)^3} \int \mathrm{d}^3\boldsymbol{x}\, \mathrm{e}^{i(\boldsymbol{k}' - \boldsymbol{k})\boldsymbol{x}} = \delta(\boldsymbol{k}' - \boldsymbol{k}). \tag{1.20}$$

After substitutions, we have

$$G(\boldsymbol{k},\omega) = -i\int \mathrm{d}t\, \mathrm{e}^{i\omega t}\left\langle T\left[c_{\boldsymbol{k}},(t)c_{\boldsymbol{k}}^{\dagger}\right]\right\rangle = \int \mathrm{d}t\, \mathrm{e}^{i\omega t}G(\boldsymbol{k},t), \tag{1.21}$$

where

$$G(\boldsymbol{k},t) = -i\left\langle T\left[c_{\boldsymbol{k}}(t),c_{\boldsymbol{k}}^{\dagger}\right]\right\rangle. \tag{1.22}$$

This shows that $G(\boldsymbol{k},\omega)$ is the time Fourier transform of $G(\boldsymbol{k},t)$. It also demonstrates the artificial origin of the factor $-i$ in the definition of the Green's function (Equation 1.10 and 1.22): it used to maintain the conventions of Fourier transforms as they were defined in Equation 1.17 and 1.21.

1.2.2 Green's function of free fermions

We can apply the formalism discussed thus far to free fermions, whose Hamiltonian is

$$H = \sum_{\boldsymbol{k}} \varepsilon_{\boldsymbol{k}} c_{\boldsymbol{k}}^{\dagger} c_{\boldsymbol{k}} \tag{1.23}$$

where $\varepsilon_{\boldsymbol{k}} = k^2/2m$. According to Equation 1.13, $c_{\boldsymbol{k}}$ has a time dependence of

$$c_{\boldsymbol{k}}(t) = \exp\left(i\varepsilon_{\boldsymbol{k}} c_{\boldsymbol{k}}^{\dagger} c_{\boldsymbol{k}} t\right) c_{\boldsymbol{k}} \exp\left(-i\varepsilon_{\boldsymbol{k}} c_{\boldsymbol{k}}^{\dagger} c_{\boldsymbol{k}} t\right), \tag{1.24}$$

which implies that it satisfies the differential equation

$$\frac{\mathrm{d}c_{\boldsymbol{k}}(t)}{\mathrm{d}t} = i\varepsilon_{\boldsymbol{k}}\mathrm{e}^{iHt}\left[c_{\boldsymbol{k}}^{\dagger}c_{\boldsymbol{k}}c_{\boldsymbol{k}} - c_{\boldsymbol{k}}c_{\boldsymbol{k}}^{\dagger}c_{\boldsymbol{k}}\right]\mathrm{e}^{-iHt} = -i\varepsilon_{\boldsymbol{k}}c_{\boldsymbol{k}}(t). \tag{1.25}$$

Assuming that at $t = 0$, $c_{\boldsymbol{k}}(t) = c_{\boldsymbol{k}}$, the solution is

$$c_{\boldsymbol{k}}(t) = \mathrm{e}^{-i\varepsilon_{\boldsymbol{k}}t}c_{\boldsymbol{k}}. \tag{1.26}$$

Substituting this result into Equation 1.22 yields

$$G(\boldsymbol{k},t) = -i\left\langle T\left[c_{\boldsymbol{k}}(t)c_{\boldsymbol{k}}^{\dagger}\right]\right\rangle = -i\left\langle \Theta(t)\mathrm{e}^{-i\varepsilon_{\boldsymbol{k}}t}c_{\boldsymbol{k}}c_{\boldsymbol{k}}^{\dagger} - \Theta(-t)\mathrm{e}^{-i\varepsilon_{\boldsymbol{k}}t}c_{\boldsymbol{k}}^{\dagger}c_{\boldsymbol{k}}\right\rangle. \tag{1.27}$$

We can identify $c_{\boldsymbol{k}}^{\dagger}c_{\boldsymbol{k}}$ as the particle number operator $n_{\boldsymbol{k}}$, whose average is equal to the particle distribution function in momentum, $f_{\boldsymbol{k}}$:

$$\left\langle c_{\boldsymbol{k}}^{\dagger}c_{\boldsymbol{k}}\right\rangle = \langle n_{\boldsymbol{k}}\rangle = f_{\boldsymbol{k}}. \tag{1.28}$$

At zero temperature, $f_{\boldsymbol{k}} = \Theta(\varepsilon_{\mathrm{F}} - \varepsilon_{\boldsymbol{k}})$. From the anti-commutation relationship, $\left\{ c_i, c_j^\dagger \right\} = \delta_{ij}$, we have

$$\left\langle c_{\boldsymbol{k}} c_{\boldsymbol{k}}^\dagger \right\rangle = \left\langle 1 - c_{\boldsymbol{k}}^\dagger c_{\boldsymbol{k}} \right\rangle = 1 - f_{\boldsymbol{k}}. \tag{1.29}$$

Substituting Equation 1.28 and 1.29 into 1.27 yields

$$G(\boldsymbol{k}, t) = -i[\Theta(t)(1 - f_{\boldsymbol{k}}) - \Theta(-t) f_{\boldsymbol{k}}] \mathrm{e}^{-i\varepsilon_{\boldsymbol{k}} t}. \tag{1.30}$$

We can calculate $G(\mathbf{k}, \omega)$ by substituting Equation 1.30 into 1.21. However, since Equation 1.30 is oscillatory because of the complex exponential term and therefore not well-defined in the limit $t \to \infty$, we take the Fourier transform of $\mathrm{e}^{-\eta|t|} G(\boldsymbol{k}, t)$ instead, with η being a positive infinitesimal number ($\eta \to 0^+$):

$$\begin{aligned} G(\boldsymbol{k}, \omega) = -\int_0^\infty \mathrm{d}t\, i(1 - f_{\boldsymbol{k}}) \mathrm{e}^{-\eta t - i\varepsilon_{\boldsymbol{k}} t + i\omega t} + \int_{-\infty}^0 \mathrm{d}t\, i f_{\boldsymbol{k}} \mathrm{e}^{\eta t - i\varepsilon_{\boldsymbol{k}} t + i\omega t} \\ = \frac{1 - f_{\boldsymbol{k}}}{\omega - \varepsilon_{\boldsymbol{k}} + i\eta} + \frac{f_{\boldsymbol{k}}}{\omega - \varepsilon_{\boldsymbol{k}} - i\eta}. \end{aligned} \tag{1.31}$$

We identify the retarded and advanced parts of the Green's function as

$$G^{\mathrm{R}}(\boldsymbol{k}, \omega) = \frac{1 - f_{\boldsymbol{k}}}{\omega - \varepsilon_{\boldsymbol{k}} + i\eta} \qquad G^{\mathrm{A}}(\boldsymbol{k}, \omega) = \frac{f_{\boldsymbol{k}}}{\omega - \varepsilon_{\boldsymbol{k}} - i\eta}. \tag{1.32}$$

A result called the Sokhotsky's formula [41] states that

$$\lim_{\eta \to 0^+} \frac{1}{\omega - \varepsilon_{\boldsymbol{k}} + i\eta} = \mathrm{P.V.} \left(\frac{1}{\omega - \varepsilon_{\boldsymbol{k}}} \right) - i\pi\delta(\omega - \varepsilon_{\boldsymbol{k}}) \tag{1.33}$$

where PV stands for Cauchy's principal value. This shows that the single-particle Green's function of free fermions has a pole at the single-particle energy of free fermions. Additionally, its imaginary part is sharply peaked (delta function) at this energy. This is the characteristic property of free particles in the language of the Green's function.

1.2.3 The spectral functions

A general expression for the Green's function can be obtained by expanding it in the basis of the Fock space, which is

$$I = |0\rangle \langle 0| + \sum_n \left| \Psi_n^1 \right\rangle \left\langle \Psi_n^1 \right| + \cdots + \sum_n \left| \Psi_n^N \right\rangle \left\langle \Psi_n^N \right| + \cdots, \tag{1.34}$$

where $|0\rangle$ represents the vacuum state and $|\Psi_n^N\rangle$ the n-th eigenstate of the Hamiltonian of an N-particle Hilbert space. Inserting the identity between $c_{\boldsymbol{k}}(t)$ and $c_{\boldsymbol{k}}^\dagger$ in Equation 1.22:

$$G(\boldsymbol{k},t) = -i\Theta(t)\sum_n \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}}(t) \middle| \Psi_n^{N+1} \right\rangle \left\langle \Psi_n^{N+1} \middle| c_{\boldsymbol{k}}^\dagger \middle| \Psi_0^N \right\rangle + i\Theta(-t)\sum_n \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}}^\dagger \middle| \Psi_n^{N-1} \right\rangle \left\langle \Psi_n^{N-1} \middle| c_{\boldsymbol{k}}(t) \middle| \Psi_0^N \right\rangle . \quad (1.35)$$

Since $|\Psi\rangle$ is the eigenstate of H,

$$\begin{aligned} \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}}(t) \middle| \Psi_n^{N+1} \right\rangle &= \left\langle \Psi_0^N \middle| e^{iHt} c_{\boldsymbol{k}} e^{-iHt} \middle| \Psi_n^{N+1} \right\rangle = e^{i\left(E_0^N - E_n^{N+1}\right)t} \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}} \middle| \Psi_n^{N+1} \right\rangle \\ \left\langle \Psi_n^{N-1} \middle| c_{\boldsymbol{k}}(t) \middle| \Psi_0^N \right\rangle &= \cdots = e^{i\left(E_n^{N-1} - E_0^N\right)t} \left\langle \Psi_n^{N-1} \middle| c_{\boldsymbol{k}} \middle| \Psi_0^N \right\rangle . \end{aligned} \quad (1.36)$$

Consequently,

$$G(\boldsymbol{k},t) = -i\Theta(t)\sum_n \exp\left[i\left(E_0^N - E_n^{N+1}\right)t\right] \left| \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}} \middle| \Psi_n^{N+1} \right\rangle \right|^2 + i\Theta(-t)\sum_n \exp\left[i\left(E_n^{N-1} - E_0^N\right)t\right] \left| \left\langle \Psi_0^N \middle| c_{\boldsymbol{k}}^\dagger \middle| \Psi_n^{N-1} \right\rangle \right|^2 . \quad (1.37)$$

As shown previously, $G(\boldsymbol{k},\omega)$ is the Fourier transform of $G(\boldsymbol{k},t)e^{-\eta|t|}$ (with $\eta \to 0^+$). Using a derivation similar to Equation 1.31, we have

$$G(\boldsymbol{k},\omega) = \sum_n \frac{\left| \left\langle \Psi_n^{N+1} \middle| c_{\boldsymbol{k}}^\dagger \middle| \Psi_0^N \right\rangle \right|^2}{\omega + E_0^N - E_n^{N+1} + i\eta} + \sum_n \frac{\left| \left\langle \Psi_n^{N-1} \middle| c_{\boldsymbol{k}} \middle| \Psi_0^N \right\rangle \right|^2}{\omega + E_n^{N-1} - E_0^N - i\eta}. \quad (1.38)$$

$\varepsilon_n^N = E_n^N - E_0^N$ can be identified as the excitation energy of an N-body system, while $\mu^N = E_0^{N+1} - E_0^N$ is the chemical potential. These definitions imply that

$$E_0^N - E_n^{N+1} = -(\varepsilon_n^{N+1} + \mu^N) \qquad E_n^{N-1} - E_0^N = \varepsilon_n^{N-1} - \mu^{N-1}. \quad (1.39)$$

When N is large, the difference between ε_n^{N+1} and ε_n^{N-1} is minute: $\varepsilon_n^{N+1} \approx \varepsilon_n^{N-1} \approx \varepsilon_n$. Similarly, $\mu^{N-1} \approx \mu$. Equation 1.38 can be re-written as

$$G(\boldsymbol{k},\omega) = \int d\omega' \left[\frac{A(\boldsymbol{k},\omega')}{\omega - \omega' - \mu + i\eta} + \frac{B(\boldsymbol{k},\omega')}{\omega + \omega' - \mu + i\eta} \right], \quad (1.40)$$

where

$$\begin{aligned} A(\boldsymbol{k},\omega') &= \sum_n \left| \left\langle \Psi_n^{N+1} \middle| c_{\boldsymbol{k}}^\dagger \middle| \Psi_0^N \right\rangle \right|^2 \delta(\varepsilon_n - \omega') \\ B(\boldsymbol{k},\omega') &= \sum_n \left| \left\langle \Psi_n^{N-1} \middle| c_{\boldsymbol{k}} \middle| \Psi_0^N \right\rangle \right|^2 \delta(\varepsilon_n - \omega'). \end{aligned} \tag{1.41}$$

$A(\boldsymbol{k},\omega')$ and $B(\boldsymbol{k},\omega')$ are known as spectral functions, which have important physical interpretations: $A(\boldsymbol{k},\omega')$ is equal to the probability of finding a particle-like excitation, while $B(\boldsymbol{k},\omega')$ corresponds to the probability of having a hole-like excitation in the N-body system. Equation 1.41 demonstrates these physical interpretations: the probabilities of finding a particle/hole-like excitation are sampled across the whole energy spectrum and the sums are the spectral functions. We have seen in Equation 1.33 that free particles have sharply defined peaks in the frequency domain. Equation 1.40 is a generalisation which remains valid when interaction is turned on. As we shall see, excitations in a Fermi liquid resemble free particles but instead of having sharply defined probability peaks, their probability distribution is spread out into broadened peaks. In essence, the spectral functions resemble probability density functions which specify the probability distribution of excitations in the momentum-energy space. Following Equation 1.33, one can show

$$\begin{aligned} \operatorname{Re} G(\boldsymbol{k},\omega) &= \text{P.V.} \int \mathrm{d}\omega' \left[\frac{A(\boldsymbol{k},\omega')}{\omega - \omega' - \mu} + \frac{B(\boldsymbol{k},\omega')}{\omega + \omega' - \mu} \right] \\ A(\boldsymbol{k},\omega) &= -\frac{1}{\pi} \operatorname{Im} G(\boldsymbol{k},\omega-\mu) \quad (\omega > \mu) \\ B(\boldsymbol{k},\omega) &= \frac{1}{\pi} \operatorname{Im} G(\boldsymbol{k},-\omega+\mu) \quad (\omega < \mu). \end{aligned} \tag{1.42}$$

1.2.4 Green's function of the Fermi liquid

We can now apply the Green's function method to the Fermi liquid model. With interaction turned on, the retarded Green's function and the spectral function becomes [61, 70]

$$\begin{aligned} G^{\mathrm{R}}(\boldsymbol{k},\omega) &= \frac{1}{\omega - \varepsilon_{\boldsymbol{k}} + i\eta - \sum(\boldsymbol{k},\omega)}, \\ A(\boldsymbol{k},\omega) &= -\frac{1}{\pi} \frac{\operatorname{Im} \sum(\boldsymbol{k},\omega)}{\left(\omega - \varepsilon_{\boldsymbol{k}} - \operatorname{Re} \sum(\boldsymbol{k},\omega)\right)^2 + \left(\operatorname{Im} \sum(\boldsymbol{k},\omega)\right)^2}, \end{aligned} \tag{1.43}$$

where $\varepsilon_{\boldsymbol{k}} = \hbar^2 k^2 / 2m + \mu$, and the additional term $\sum(\boldsymbol{k},\omega)$ is known as the self-energy. It represents the contribution to a quasiparticle's energy due to interactions between itself and its surroundings. We see that with a non-zero self-energy, the spectral function is a Lorentzian peak centred at $\omega = \varepsilon_{\boldsymbol{k}} + \operatorname{Re} \sum(\boldsymbol{k},\omega)$, with width equal to the imaginary part

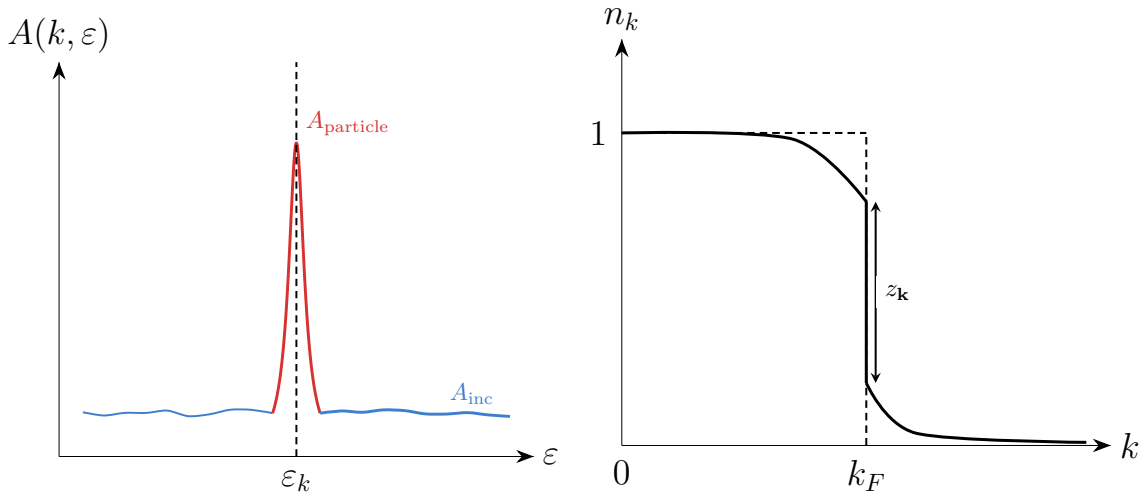


FIGURE 1.2 Left: spectral function of the Fermi liquid, consisting of a Lorentzian quasiparticle peak of weight $z_{\boldsymbol{k}}$ and an incoherent background. Right: the momentum distribution of bare particles in a Fermi liquid at 0 temperature. Dashed profile: momentum distribution of bare electrons without interactions. Solid profile: distribution of bare electrons with interactions. A gap of size $z_{\boldsymbol{k}}$ across the Fermi surface is a characteristic of the Fermi liquid. n_k represents the probability of occupancy of a state with momentum $\hbar k$.

of the self-energy, $\text{Im}\sum(\boldsymbol{k},\omega)$. The centre of the spectral function defines the energy of the quasiparticle; the width of the spectral function determines the life time of the quasiparticle. The second statement can be proved by considering the Green's function in the time domain, modified by an exponential decay term to account for a finite life-time λ:

$$G^{\text{R}}(\boldsymbol{k},t) = -i\Theta(t)\exp\left(-i\varepsilon_{\boldsymbol{k}}t\right)\exp\left(-\frac{t}{\lambda}\right). \tag{1.44}$$

Its Fourier transform and spectral function are

$$G^{\text{R}}(\boldsymbol{k},\omega) = \frac{1}{\omega-\varepsilon_{\boldsymbol{k}}+i\eta+i/\lambda} \qquad A(\boldsymbol{k},\omega) = \frac{1}{\pi}\frac{1/\lambda}{(\omega-\varepsilon_{\boldsymbol{k}})^2+(1/\lambda)^2}. \tag{1.45}$$

One can see that the spectral function width corresponds directly to the inverse of the life-time of the quasiparticle (also known as the scattering rate).

The introduction of self-energy to the spectral function also results in a change to the area of the peak (referred to as the spectral weight). For free particles, the delta function spectral peak has unit area; for quasiparticles, one can expand the real term of the spectral function around the quasiparticle peak as

$$\omega-\varepsilon_{\boldsymbol{k}}-\text{Re}\sum(\boldsymbol{k},\omega) \simeq 0+(\omega-\varepsilon_{\boldsymbol{k}})\frac{\partial}{\partial\omega}\left[\omega-\varepsilon_{\boldsymbol{k}}-\text{Re}\sum(\boldsymbol{k},\omega)\right]\Big|_{\omega=\varepsilon_{\boldsymbol{k}}} = \frac{\omega-\varepsilon_{\boldsymbol{k}}}{z_{\boldsymbol{k}}}, \tag{1.46}$$

which prompts us to rewrite the quasiparticle spectral function as

$$A(\boldsymbol{k},\omega) = \frac{z_{\boldsymbol{k}}}{\pi}\frac{1/\tau}{(\omega-\varepsilon_{\boldsymbol{k}})^2+(1/\tau)^2} + A_{\text{inc}}, \tag{1.47}$$

where the first term is a Lorentzian peak of spectral weight $z_{\boldsymbol{k}}$. Due to the energy dependence of the self-energy, $z_{\boldsymbol{k}} = \partial_\omega^{-1}[\omega - \varepsilon_{\boldsymbol{k}} - \operatorname{Re}\sum(\boldsymbol{k},\omega)] = 1\big/(1-\partial_\omega \operatorname{Re}\sum(\boldsymbol{k},\omega))$ is usually a number smaller than 1. The extra term A_{inc} is required to account for the rest of the spectral weight $(1 - z_{\boldsymbol{k}})$. The Lorentzian peak corresponds to the quasiparticles. Excitation in this part of the spectrum resembles real particles with well-defined dispersion relation $\omega = \varepsilon_{\boldsymbol{k}} + \operatorname{Re}\sum(\boldsymbol{k},\omega)$ and finite life-time τ. The rest of the spectrum is filled with a broad, featureless background A_{inc} which corresponds to the incoherent excitation. As interaction is switched on, the collective motion turns a portion $z_{\boldsymbol{k}}$ of electrons into quasiparticles. As shown in Figure 1.2, this causes a gap to open up in the momentum distribution of free particles across the Fermi level. Essentially, interaction causes a portion of the originally occupied states to be promoted above the Fermi energy as particle-like excitations, leaving behind hole-like empty states. The proportion of states that undergo this re-normalisation is usually small, meaning that the new particle distribution still resembles the original free-particle distribution (which is a delta function). We can still identify the system as consisting of mostly particles below the Fermi surface and empty states above it. The re-distributed states can be considered as quasiparticles which rarely interact with the bare particles or each other.

1.2.5 Finite temperature and analytic continuation

The purely quantum mechanical formalism discussed up to this point applies only when the system temperature is zero. In order to account for the effect of finite temperature, the formalism needs to incorporate statistical mechanics: the expectation value of a generic observable A at finite temperature T is given by

$$\langle A\rangle = \frac{\langle\Psi_0|\mathrm{e}^{-\beta K}A|\Psi_0\rangle}{\langle\Psi_0|\mathrm{e}^{-\beta K}|\Psi_0\rangle}, \tag{1.48}$$

where $|\Psi_0\rangle$ is the ground state of the many-body system; $\beta = 1\big/k_{\mathrm{B}}T$ is the inverse temperature; $K = H - \mu N$ is the grand-canonical Hamiltonian, with μ being the chemical potential and N the number of particles. In field theory this average is evaluated by the following functional integral:

$$\langle A(r)\rangle = \frac{\int \mathcal{D}\psi^\dagger \mathcal{D}\psi\ \mathrm{e}^{-S[\psi^\dagger,\psi]}A(r)}{\int \mathcal{D}\psi^\dagger \mathcal{D}\psi\ \mathrm{e}^{-S[\psi^\dagger,\psi]}}, \tag{1.49}$$

where $r = (x, v\tau)$ is known as the general space-time coordinate, with $\tau = ti$ being the imaginary time, and v the speed of a particle; $\int \mathcal{D}\psi$ represents the functional integral over all possible functions of $\psi(r)$, the field operator; S is the action, which is related to the Hamiltonian via: $S = \int \mathrm{d}t\,(-\dot{\boldsymbol{r}} \cdot \boldsymbol{p} - H)$, with $\boldsymbol{p}$ being the momentum operator.

The use of imaginary time allows theorists to apply a mathematical technique known as analytic continuation: essentially, results obtained from the finite temperature formalism can be shown to converge with that of the zero-temperature formalism in the limit $T \to 0$. This is because exponential factors such as $\exp\left[i\left(E_0^N - E_n^{N+1}\right)t\right]$ in Equation 1.37 and $\mathrm{e}^{-\beta K}$ in Equation 1.48 converge on each other at $T = 0$ when one makes the substitution of imaginary time. It is worth noting that in terms of imaginary time, instead of describing the Green's function as consisting of a retarded and advanced part, one would refer to the parts as the particle branch and hole branch. (This can be seen from the numerators of Equation 1.32)

1.3 The Luttinger liquid

Although the Fermi-liquid theory has been successfully applied to match the experimental data of many fermionic systems, from the isotropic liquid ^{3}He to conduction electrons in metals and semiconductors, there exist exotic systems in which the theory breaks down. These exceptions are collectively described as non-Fermi liquids and occur when the fundamental excitations do not resemble bare particles like the quasiparticles (a review of these systems can be found in [78]). One instance of special interest to us is the case of a fermion chain constrained in one dimension. Unlike non-Fermi liquids in higher dimensions, for which only approximate models exist, a new type of adiabatic continuity and an explicitly solvable model has been proposed by Haldane [36], known as the Tomonaga-Luttinger liquid (TLL), to quantify the state of 1D fermions.

While it is remarkable that a novel physical model can emerge from simply constraining a system into a lower dimension, one can intuitively see how the dynamics of interaction is drastically changed in 1D: in higher dimensions, an electron can propagate through the Fermi sea by pushing away its neighbours, and the excitations generated this way resembles a particle dressed by a cloud; in contrast, an electron in a 1D chain cannot move past its neighbours, hence any excitation along the chain must result in a collective response of the entire system. One can treat the fundamental excitation of an 1D electron gas as a particle-hole pair (Figure 1.3), where a particle of momentum k is destroyed below the Fermi energy while another of momentum $k + q$ is created above ($\hbar$ is suppressed in this discussion so that wavenumbers have the dimension of momentum). The momentum of this excitation is exactly q. In higher dimensions, since the vector $\boldsymbol{q}$ can be rotated freely, the energy of a particle-hole excitation

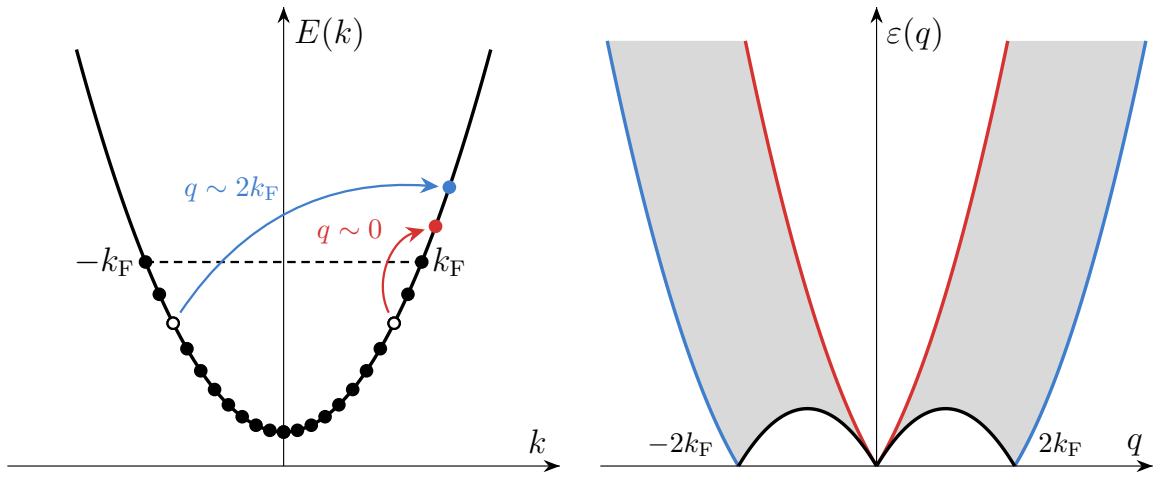


FIGURE 1.3 Left: single-particle dispersion of free 1D fermions with small (red) and large (blue) energy particle-hole excitations across the Fermi point. The filled dots represent the occupied states and the empty ones the holes. Right: energy spectrum (indicated by the grey area) of single 1D electron-hole pairs. The excitations in the left plot coincide with the same-coloured branch on the energy spectrum. The forbidden region indicated by the gap is unique to 1D: in higher dimensions these energies are accessible. The black curve that marks the boundary between the accessible and forbidden regions is the spectral edge.

can take on any value; in 1D however, as $\boldsymbol{q}$ does not possess any angular freedom, a minimum energy must be spent to remove a particle along the dispersion, leading to a gap in the excitation spectrum. For example, only with $q = 0$ or $2k_\mathrm{F}$ can the excitation energy be zero. Along this minimum energy dispersion, one may expand the energy around the Fermi points to get

$$E(k) = v_\mathrm{F}(k - k_\mathrm{F}) + \mathcal{O}\left((k - k_\mathrm{F})^2\right). \tag{1.50}$$

From this relation one sees that the excitation energy is

$$\varepsilon(q) = E(k + q) - E(k) = v_\mathrm{F} q + \mathcal{O}\left((k - k_\mathrm{F})^2\right). \tag{1.51}$$

Additionally, the differential of the energy is

$$\delta\varepsilon(q) = \mathcal{O}(q^2, k^2). \tag{1.52}$$

The two results above show that particle-hole pairs have well-defined momentum and energy regardless of the form of the energy dispersion (e. g. the dispersion can be that of a free electron, or a band structure resulting from a lattice). More importantly, the energy dispersion vanishes faster than the energy itself when $\varepsilon(q) \to 0$, implying that they are stable, long-lived excitations when close to the Fermi energy—particle-hole pairs in 1D take on a role similar

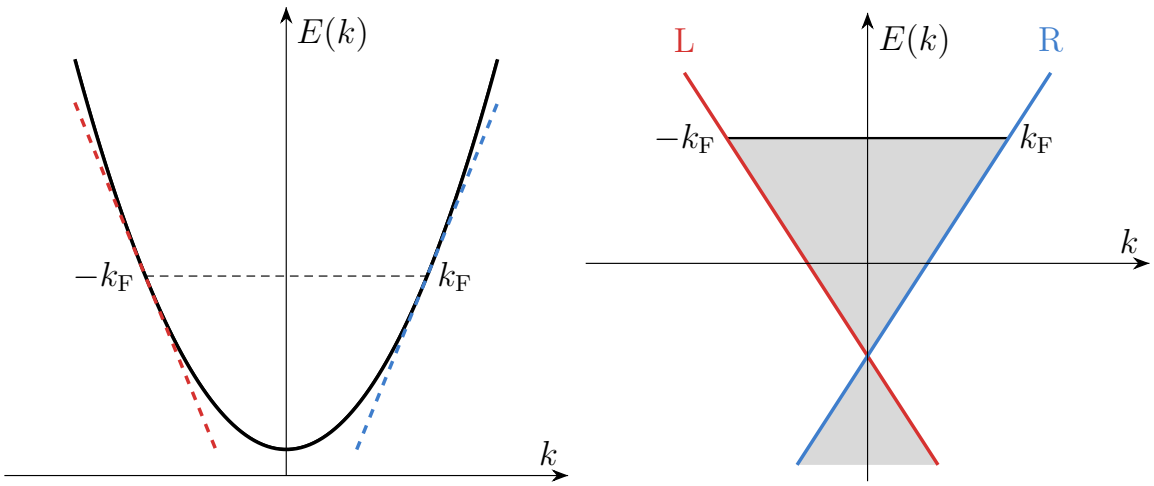


FIGURE 1.4 The Tomonaga-Luttinger model for 1D fermions. Left: the original fermion energy dispersion relation. The dashed lines show that the dispersion can be approximated by a linear function close to the Fermi points. Right: a linear dispersion with states below the Fermi energy (including all negative energies) occupied. The two branches correspond to the left-moving and right-moving fermions, respectively. The shaded area is the Dirac sea. This dispersion is used in the TLL model.

to that of quasiparticles in higher dimensions. Since a particle-hole pair involves the creation and annihilation of an electron, it should be bosonic in nature. Summing up the arguments made thus far, one would expect the excitations in 1D to be described in a bosonic basis of collective motion.

1.3.1 The spinless Tomonaga-Luttinger model

An exact solution to the many-body problem of 1D interacting fermions is given by the Tomonaga-Luttinger model, which uses a technique known as bosonisation. Equation 1.50 showed that the energy dispersion near the Fermi points can be approximated by a linear expression. This linear approximation is taken to the extreme in the TLL model, where the original energy spectrum is replaced by a linear one which is occupied all the way down to $-\infty$ on the energy scale (Figure 1.4). The occupied states make up what is known as the Dirac sea. Referring to Equation 1.50, the Hamiltonian of free 1D fermions (also referred to as the kinetic Hamiltonian) can be written with the single-particle operators as

$$H_0 = \sum_{k,r} v_{\mathrm{F}}(\epsilon_r k - k_{\mathrm{F}}) c^{\dagger}_{r,k} c_{r,k}, \tag{1.53}$$

where $r = \mathrm{L}$ or R labels the left- or right-moving fermions, $\epsilon_{\mathrm{R}} = 1$ and $\epsilon_{\mathrm{L}} = -1$. As discussed previously, 1D excitations should be described in the basis of particle-hole pairs. The quantity

that directly corresponds to them is the density fluctuation, given by the operator

$$\rho_r^\dagger(p) = \sum_k c_{r,k+p}^\dagger c_{r,k}. \tag{1.54}$$

Because of the inclusion of all negative states in the model, the density operator defined above is infinite. To allow mathematical manipulation we introduce a finite quantity known as the normal ordering of operators (denoted by double colon), which is equivalent to an operator minus its ground-state (the Dirac sea) average:

$$: \rho_r^\dagger(p) := \sum_k c_{r,k+p}^\dagger c_{r,k} - \langle 0|c_{r,k+p}^\dagger c_{r,k}|0\rangle\,. \tag{1.55}$$

Note that one can identify the normal ordering of density as the number operator for each species of moving fermions when $p = 0$:

$$: \rho_r^\dagger(0) := \sum_k c_{r,k}^\dagger c_{r,k} - \langle 0|c_{r,k}^\dagger c_{r,k}|0\rangle = N_r. \tag{1.56}$$

One can calculate the commutation relation of the density operator for fermions moving in the same direction:

$$\begin{aligned}
\left[\rho_r^\dagger(p), \rho_r^\dagger(-p')\right] &= \sum_{k_1,k_2} \left[c_{r,k_1+p}^\dagger c_{r,k_1}, c_{r,k_2-p'}^\dagger c_{r,k_2}\right] \\
&= \sum_{k_1,k_2} c_{r,k_1+p}^\dagger c_{r,k_2}\delta_{k_1,k_2-p'} - c_{r,k_2-p'}^\dagger c_{r,k_1}\delta_{k_1+p,k_2} = \sum_{k_2} c_{r,k_2-p'+p}^\dagger c_{r,k_2} - c_{r,k_2-p'}^\dagger c_{r,k_2-p} \\
&= \sum_{k_2} : c_{r,k_2-p'+p}^\dagger c_{r,k_2} : - : c_{r,k_2-p'}^\dagger c_{r,k_2-p} : + \langle 0|c_{r,k_2-p'+p}^\dagger c_{r,k_2}|0\rangle - \langle 0|c_{r,k_2-p'}^\dagger c_{r,k_2-p}|0\rangle \\
&= \sum_{k_2} \langle 0|c_{r,k_2-p'+p}^\dagger c_{r,k_2}|0\rangle - \langle 0|c_{r,k_2-p'}^\dagger c_{r,k_2-p}|0\rangle = \delta_{p,p'} \sum_{k_2} \langle 0|c_{r,k_2}^\dagger c_{r,k_2}|0\rangle - \langle 0|c_{r,k_2-p}^\dagger c_{r,k_2-p}|0\rangle\,.
\end{aligned} \tag{1.57}$$

- To derive the second equality, one manipulates the commutator

$$\left[c_1^\dagger c_2, c_3^\dagger c_4\right] = c_1^\dagger c_2 c_3^\dagger c_4 - c_3^\dagger c_4 c_1^\dagger c_2 \tag{†}$$

by repeatedly swapping pairs of operators in the second term until the indices are ordered from 1 to 4 from left to right. Applying the fermion anti-commutation relation, $\left\{c_i^\dagger, c_j\right\} =$

δ_{ij} and $\left\{c_i^\dagger, c_j^\dagger\right\} = \{c_i, c_j\} = 0$, we have $c_i^\dagger c_j = \delta_{ij} - c_j c_i^\dagger$ and $c_i^\dagger c_j^\dagger = c_j^\dagger c_i^\dagger$. Therefore,

$$\begin{aligned} c_3^\dagger c_4 c_1^\dagger c_2 &= c_3^\dagger(\delta_{1,4} - c_1^\dagger c_4)c_2 = c_3^\dagger \delta_{1,4} c_2 - c_3^\dagger c_1^\dagger c_4 c_2 \\ &= c_3^\dagger \delta_{1,4} c_2 - c_1^\dagger c_3^\dagger c_2 c_4 = c_3^\dagger \delta_{1,4} c_2 - c_1^\dagger(\delta_{2,3} - c_2 c_3^\dagger)c_4 \\ &= c_3^\dagger \delta_{1,4} c_2 - c_1^\dagger \delta_{2,3} c_4 + c_1^\dagger c_2 c_3^\dagger c_4. \end{aligned} \qquad (\dagger\dagger)$$

The δ terms are left after †† is substituted back into †.

- In the third line, the two normal-ordered products can be shown to cancel by making a change of variable $k_2 + p \to k_2$ in the first term.

The difference between the two ground-state averages in the final expression of Equation 1.57 is the difference in the number of states between two Dirac sea, offset by $-p$ units of momentum. For states quantised according to the periodic boundary condition, $k_n = 2\pi n/L$, this gives rise to

$$\left[\rho_r^\dagger(p), \rho_r^\dagger(-p')\right] = -\epsilon_r \delta_{p,p'} \frac{pL}{2\pi}, \qquad (1.58)$$

where L is the length of the chain. This commutation relation demonstrates the bosonic nature of the density operator, as it has the form of the boson commutation relation, $[b_i^\dagger, b_j] = \delta_{ij}$. One can define normalised bosonic creation and annihilation operators for 1D excitations as follows:

$$b_p^\dagger = \sqrt{\frac{2\pi}{|p|L}} \sum_r \Theta(\epsilon_r p)\rho_r^\dagger(p) \qquad b_p = \sqrt{\frac{2\pi}{|p|L}} \sum_r \Theta(\epsilon_r p)\rho_r^\dagger(-p), \qquad (1.59)$$

where Θ is the step function. Note that these operators are defined for $p \neq 0$. By expanding these operators in terms of $c^\dagger$ and c, it can be shown that they satisfy the following commutators with the kinetic Hamiltonian defined in Equation 1.53:

$$\left[H_0, b_p^\dagger\right] = v_\mathrm{F}|p|b_p^\dagger \qquad \left[H_0, b_p\right] = -v_\mathrm{F}|p|b_p, \qquad (1.60)$$

which resembles the commutation relation between a harmonic oscillator Hamiltonian and its ladder operators. The Hamiltonian in terms of these bosonic operators can be written as (see Chapter 2 of [27], and [90])

$$H_0 = \sum_{p\neq 0} v_\mathrm{F}|p|b_p^\dagger b_p + \frac{v_\mathrm{F}\pi}{L} \sum_r N_r^2. \qquad (1.61)$$

Note that the first term satisfies Equation 1.60. The second term is equal to the limit of the first term as $p \to 0$. This expression shows that the kinetic Hamiltonian remains a quadratic function when one switches from the basis of the fermionic operators to the bosonic operators.

The density and current fields

Instead of the bosonic operators introduced in the last section, the Hamiltonian H_0 was historically expressed in terms of two fields, $\phi(x)$ and $\theta(x)$, which are related to the boson operators and have special physical interpretations that will be discussed shortly. The fields are

$$\begin{aligned}\phi(x) &= -(N_{\mathrm{R}} + N_{\mathrm{L}})\frac{\pi x}{L} - \frac{i\pi}{L}\lim_{\alpha\to 0}\sum_{p\neq 0}\frac{1}{p}\exp\left(-\frac{\alpha|p|}{2} - ipx\right)(\rho_{\mathrm{R}}^{\dagger}(p) + \rho_{\mathrm{L}}^{\dagger}(p)) \\ \theta(x) &= (N_{\mathrm{R}} - N_{\mathrm{L}})\frac{\pi x}{L} + \frac{i\pi}{L}\lim_{\alpha\to 0}\sum_{p\neq 0}\frac{1}{p}\exp\left(-\frac{\alpha|p|}{2} - ipx\right)(\rho_{\mathrm{R}}^{\dagger}(p) - \rho_{\mathrm{L}}^{\dagger}(p)).\end{aligned} \tag{1.62}$$

And the free Hamiltonian is

$$H_0 = \frac{1}{2\pi}\int \mathrm{d}x\, v_{\mathrm{F}}\left[(\nabla\theta)^2 + (\nabla\phi)^2\right]. \tag{1.63}$$

The three formulae above can be derived from attempts to express single-particle operators using the bosonic operator $\rho^{\dagger}(p)$. Suppose $\psi_r(x)$ is the field operator which destroys an r-moving electron at position x. It is related to the annihilation operator in momentum space, $c_{r,k}$, via the relation [82]

$$\psi_r(x) = \frac{1}{\sqrt{L}}\sum_k \mathrm{e}^{ikx}c_{r,k}. \tag{1.64}$$

Its commutation relation with the density operator is

$$\left[\rho_r^{\dagger}(p), \psi_r(x)\right] = \frac{1}{\sqrt{L}}\sum_{k,k'}\mathrm{e}^{ik'x}\left[c_{r,k+p}^{\dagger}c_{r,k}, c_{r,k'}\right] = -\mathrm{e}^{ipx}\psi_r(x). \tag{1.65}$$

A function that satisfies the commutation relation is

$$\psi_r'(x) = \exp\left\{\sum_p \mathrm{e}^{ipx}\rho_r^{\dagger}(-p)\left(\frac{2\pi\epsilon_r}{pL}\right)\right\}, \tag{1.66}$$

which can be verified by noting that for any two operators A and B, $[A, \mathrm{e}^B] = \sum_{i=0}^{\infty} \frac{[A,B^i]}{i!}$. However, $\psi'(x)$ cannot be the correct operator: while one expects $\psi(x)$ to reduce the number of fermions by 1, $\psi'(x)$ cannot change the number of fermions in the system because it is purely a function of the bosonic density operator, which conserves the number of fermions. It is therefore necessary to introduce extra operators U_r, known as the Klein factors, to the definition of $\psi(x)$

$$\psi_r(x) = U_r \exp\left\{ \sum_p \mathrm{e}^{ipx} \rho_r^\dagger(-p) \left(\frac{2\pi\epsilon_r}{pL} \right) \right\}, \tag{1.67}$$

where U_r and its complex conjugate satisfy in the case $r = R$:

$$U_\mathrm{R} \left| N_\mathrm{R}, N_\mathrm{L} \right\rangle = \left| N_\mathrm{R} - 1, N_\mathrm{L} \right\rangle \qquad U_\mathrm{R}^\dagger \left| N_\mathrm{R}, N_\mathrm{L} \right\rangle = \left| N_\mathrm{R} + 1, N_\mathrm{L} \right\rangle, \tag{1.68}$$

and similarly for $r = \mathrm{L}$. The function of U_r is to ensure that $\psi(x)$ observes the proper fermionic commutation order. It turns out that such an operator has the following form:

$$U_r = \frac{1}{\sqrt{L}} \int_0^L \mathrm{d}x\, \mathrm{e}^{-i\epsilon_r k_\mathrm{F}} \mathrm{e}^{i\phi_r(x)} \psi_r(x) \mathrm{e}^{i\phi_r^\dagger(x)}, \tag{1.69}$$

where the field

$$\phi_r(x) = -\frac{\epsilon_r \pi x}{L} N_r + \lim_{\epsilon\to 0} i \sum_{p\neq 0} \sqrt{\frac{2\pi}{L|p|}} \exp\left(-\frac{\epsilon L|p|}{2\pi} \right) \mathrm{e}^{ipx} \Theta(\epsilon_r p) b_p \tag{1.70}$$

makes U_r commute with the boson operator. While the Klein factor is required to produce a rigorous mapping between the bosonic field and the fermionic excitations, it is not needed for the calculations of correlation functions. We therefore refer to Appendix B of [27] for more detailed derivations. From Equation 1.69 one can deduce the single-particle operator

$$\begin{aligned} \psi_r(x) &= \frac{1}{\sqrt{L}} \mathrm{e}^{i\epsilon_r k_\mathrm{F}} \mathrm{e}^{-i\phi_r(x)} U_r \mathrm{e}^{-i\phi_r^\dagger(x)} \\ &= \lim_{\epsilon\to 0} \frac{1}{\sqrt{2\epsilon L}} \exp\left[i\epsilon_r \left(k_\mathrm{F} - \frac{\pi}{L} \right) x \right] \exp\left[-i\phi_r(x) - i\phi_r^\dagger(x) \right] U_r \\ &= \lim_{\alpha\to 0} \frac{1}{\sqrt{2\pi\alpha}} \exp\left[i\epsilon_r \left(k_\mathrm{F} - \frac{\pi}{L} \right) x \right] \exp\left[-i\epsilon_r \phi(x) + i\theta(x) \right] U_r. \end{aligned} \tag{1.71}$$

In the last line the bosonic fields $\phi(x)$ and $\theta(x)$ arise because of the Klein factor. Referring back to Equation 1.62, we see that the effect of the limit $\alpha = \epsilon L/\pi \to 0$ is to ensure that the fields have a finite integral over the whole x space (similar to a slowly decaying oscillating

wavefunction). Having worked out $\psi(x)$, the linearised free Hamiltonian can be derived as: $H_0 = \sum_r \int \mathrm{d}x\, \psi_r^\dagger(x)[v_\mathrm{F}(-i\nabla - k_\mathrm{F})]\psi_r(x)$, where $-i\nabla$ is the momentum operator.

The significance of the two fields can be deduced from Equation 1.62. Firstly, in the thermal dynamic limit ($L \to \infty$, i. e. the system size is macroscopic)

$$\nabla\phi = -\pi[\rho_\mathrm{R}(x) + \rho_\mathrm{L}(x)] \qquad \nabla\theta = \pi[\rho_\mathrm{R}(x) - \rho_\mathrm{L}(x)], \tag{1.72}$$

where $\rho_r(x)$ is the density operator in position representation

$$\rho_r(x) = \frac{1}{L}\sum_p \mathrm{e}^{-ikx}\rho_r^\dagger(p). \tag{1.73}$$

Being proportional to the sum of the density operators of the left- and right-movers, $\nabla\phi(x)$ can be identified as the total density variation at position x; being proportional to their difference, $\nabla\theta(x)$ corresponds to the net current along the chain. Secondly,

$$\begin{aligned}\big[\phi(x_1), \theta(x_2)\big] = \lim_{\alpha\to 0}\sum_{p\neq 0}\frac{\pi}{Lp}\exp\big[ip(x_2 - x_1) - \alpha|p|\big] &= \int_0^\infty \frac{\mathrm{d}p}{p}\exp\big[ip(x_2 - x_1)\big] \\ &= i\int_0^\infty \mathrm{d}p\,\frac{\sin[p(x_2 - x_1)]}{p} = i\frac{\pi}{2}\,\mathrm{Sign}(x_2 - x_1),\end{aligned} \tag{1.74}$$

where the last equality is known as the Dirichlet integral. In general, for any two operators that satisfy $[A(x), B(y)] = C(x, y)$, one has $[\nabla A, B] = \nabla C$. Therefore,

$$\big[\phi(x_1), \nabla\theta(x_2)\big] = i\pi\delta(x_2 - x_1). \tag{1.75}$$

This result is significant because it means that $\phi(x)$ and $\nabla\theta(x)\big/\pi$ satisfy the canonical commutation relation (sometimes a substitutional field $\Pi(x) = \nabla\theta(x)\big/\pi$ is introduced to simplify the expression). In essence, the Hamiltonian of Equation 1.63 resembles that of a harmonic oscillator, with $\phi(x)$ taking on the role of the position operator and $\nabla\theta(x)\big/\pi$ (or $\Pi(x)$) the momentum operator. This demonstrates one of the advantages of the bosonisation method: by making the change of basis to the bosonic fields, the kinetic Hamiltonian assumes a familiar, exactly solvable form.

The effect of interactions

The Hamiltonian of interactions can be constructed from the total density operators as

$$H_\mathrm{int} = \int \mathrm{d}x\,\mathrm{d}x'\, V(x - x')\rho(x)\rho(x'), \tag{1.76}$$

where $\rho(x)$ is related to the single-particle operators by $\rho(x) = \psi^\dagger(x)\psi(x)$, with

$$\psi(x) = \frac{1}{\sqrt{L}} \sum_r \left[\sum_{k-k_F=-\Lambda}^{\Lambda} e^{ikx} c_{r,k} + \sum_{k+k_F=-\Lambda}^{\Lambda} e^{ikx} c_{r,k} \right] = \psi_L(x) + \psi_R(x), \tag{1.77}$$

where $\psi_r(x)$ is defined in Equation 1.64. Since we are only concerned with a narrow band of momentum around the two Fermi points where the linear dispersion is a good approximation, Λ represents the small momentum cutoff of the model. The total density operator can be expanded as

$$\begin{aligned} \rho(x) = \psi^\dagger(x)\psi(x) &= \left(\psi_L^\dagger(x) + \psi_R^\dagger(x)\right)\left(\psi_L(x) + \psi_R(x)\right) \\ &= \left[\psi_L^\dagger(x)\psi_L(x) + \psi_R^\dagger(x)\psi_R(x)\right] + \left[\psi_L^\dagger(x)\psi_R(x) + \psi_R^\dagger(x)\psi_L(x)\right]. \end{aligned} \tag{1.78}$$

The result can be organised into two groups as was indicated by the brackets. The first group corresponds to scattering between fermions moving in the same direction (which lead to excitations with $q \sim 0$); the second group corresponds to scattering between opposite-moving fermions (excitations with $q \sim 2k_F$). Using this result, we see that the interaction Hamiltonian, which consists of the product of two density operators, must expand into a quartic function in terms of the fermion operators. Such complexity of the interaction Hamiltonian is what makes the fermion operator basis ill-suited to the description of excitations: one cannot easily diagonalise a quartic Hamiltonian. We will find out below that once we make the change of basis into the bosonic operators, the interaction Hamiltonian will reduce to a quadratic form, and therein lies the exact solution of the model.

If we consider the interaction to be short-ranged and the excitation to be a continuum (model described in Figure 1.3), then we only need to consider the interaction between low-level ($q \sim 0$) excitations [87]. We would have, according to Equation 1.78, $\rho(x) = \rho_L(x) + \rho_R(x)$, where $\rho_r(x) = \psi_r^\dagger(x)\psi_r(x)$. For short-range interactions, one can assume $V(x-x') = V_0\delta(x-x')$, where V_0 is a constant which characterises the interaction strength. Substituting these back into the Hamiltonian of Equation 1.76 gives

$$H_{\text{int}} = V_0 \int \mathrm{d}x\, \rho(x)\rho(x) = V_0 \int \mathrm{d}x \left[\left[\rho_L(x)\right]^2 + \left[\rho_R(x)\right]^2 + 2\rho_L(x)\rho_R(x) \right]. \tag{1.79}$$

Rearranging Equation 1.72, one has

$$\rho_R(x) = (\nabla\theta - \nabla\phi)/2\pi \qquad \rho_L(x) = -(\nabla\theta + \nabla\phi)/2\pi. \tag{1.80}$$

Applying these results to Equation 1.79 leads to

$$H_{\text{int}} = V_0 \int \mathrm{d}x \left(\frac{\nabla\phi + \nabla\theta}{2\pi}\right)^2 + \left(\frac{\nabla\phi - \nabla\theta}{2\pi}\right)^2 + \frac{2\,(\nabla\phi)^2 - 2\,(\nabla\theta)^2}{(2\pi)^2} = \frac{4V_0}{(2\pi)^2} \int \mathrm{d}x\,(\nabla\phi)^2. \quad (1.81)$$

We can now write the total Hamiltonian of the TLL using Equations 1.63 and 1.81 as

$$H_{\text{TLL}} = H_0 + H_{\text{int}} = \frac{v_{\text{F}}}{2\pi} \int \mathrm{d}x\,[(\nabla\theta)^2 + (\nabla\phi)^2] + \frac{4V_0}{(2\pi)^2} \int \mathrm{d}x\,(\nabla\phi)^2 = \frac{v_{\text{F}}}{2\pi} \int \mathrm{d}x\,(\nabla\theta)^2 + \left(\frac{v_{\text{F}}}{2\pi} + \frac{4V_0}{(2\pi)^2}\right) \int \mathrm{d}x\,(\nabla\phi)^2. \quad (1.82)$$

This result can be simplified by introducing the Luttinger parameters v (with dimension of velocity) and K (dimensionless)

$$v = v_{\text{F}}\sqrt{1 + \frac{2V_0}{\pi v_{\text{F}}}} \qquad K = \frac{1}{\sqrt{1 + \frac{2V_0}{\pi v_{\text{F}}}}}. \quad (1.83)$$

The parameters satisfy the relationships

$$vK = v_{\text{F}} \qquad \frac{v}{K} = v_{\text{F}}\left(1 + \frac{2V_0}{\pi v_{\text{F}}}\right), \quad (1.84)$$

under which the Hamiltonian can be written as

$$H_{\text{TLL}} = \frac{v}{2\pi} \int \mathrm{d}x \left[K(\nabla\theta)^2 + \frac{1}{K}(\nabla\phi)^2\right]. \quad (1.85)$$

Equation 1.85 is the Hamiltonian of the Luttinger model that is regularly quoted in literature. We see that in the basis of the bosonic operators, interaction does not change the fundamental form of the Hamiltonian. Most importantly it remains quadratic and consequently exactly solvable. In deriving the Hamiltonian, we have assumed a particular interaction, which was the δ-function. While the exact expression of the Luttinger parameters in terms of the microscopic interaction parameters will be different with a different choice of interaction function, the total Hamiltonian will retain the form of Equation 1.85 regardless of the choice (see Chapter 2, [27] for a generalised proof using Feynman diagrams). Additionally, the relationship $vK = v_F$ remains valid for systems with Galilean invariance [65].

In general, a repulsive interaction will result in $K < 1$ and an attractive one will result in $K > 1$.

1.3.2 The spinful model

Up to now we have ignored the spin label of fermions in order to reduce the complexity of the bosonisation process. The results obtained thus are categorically referred to as the 'spinless' TLL model in literature. Having studied the simplified model, we can now consider the 'spinful' model in which the spin labels are accounted for.

With the introduction of the spin label $\sigma =\uparrow$ or $\downarrow$, the kinetic Hamiltonian becomes

$$H_0 = \sum_{k,r,\sigma} v_{\mathrm{F}}(\epsilon_r k - k_{\mathrm{F}}) c^{\dagger}_{r,k,\sigma} c_{r,k,\sigma}. \tag{1.86}$$

Since there are no coupling terms between opposite spins, the Hamiltonian represented by the bosonic fields is simply

$$H_0 = \frac{v_{\mathrm{F}}}{2\pi} \sum_{\sigma} \int \mathrm{d}x \, [(\nabla\theta_\sigma)^2 + (\nabla\phi_\sigma)^2]. \tag{1.87}$$

The interacting Hamiltonian contains a product between two density operators $\rho(x)$, which with the inclusion of spin becomes $\rho(x) = \rho_\uparrow(x) + \rho_\downarrow(x)$. Assuming the same short-range δ-function interaction potential as before,

$$H_{\mathrm{int}} = V_0 \int \mathrm{d}x \, (\rho(x))^2 = V_0 \int \mathrm{d}x \, (\rho_\uparrow^2 + \rho_\downarrow^2 + 2\rho_\uparrow\rho_\downarrow), \tag{1.88}$$

where we have combined the two cross-terms because opposite spins are independent and therefore their labels commute. The density operator of each spin can be split into a left- and right-moving branch: $\rho_\sigma = \rho_{\mathrm{L},\sigma} + \rho_{\mathrm{R},\sigma}$. We can add the spin labels to Equation 1.80 to obtain

$$\rho_{\mathrm{R},\sigma} = \frac{1}{2\pi}(\nabla\theta_\sigma - \nabla\phi_\sigma) \qquad \rho_{\mathrm{L},\sigma} = -\frac{1}{2\pi}(\nabla\theta_\sigma - \nabla\phi_\sigma). \tag{1.89}$$

Consequently,

$$\begin{aligned} \rho_\uparrow = \rho_{\mathrm{R},\uparrow} + \rho_{\mathrm{L},\uparrow} = \frac{1}{2\pi}(\nabla\theta_\uparrow - \nabla\phi_\uparrow - \nabla\theta_\uparrow + \nabla\phi_\uparrow) = \frac{2}{2\pi}\nabla\phi_\uparrow \\ \rho_\downarrow = \cdots = \frac{2}{2\pi}\nabla\phi_\downarrow. \end{aligned} \tag{1.90}$$

Substituting these results back into Equation 1.88, we have

$$H_{\text{int}} = \sum_{\sigma} \frac{4V_0}{(2\pi)^2} \int \mathrm{d}x\, (\nabla\phi_\sigma)^2 + \frac{8V_0}{(2\pi)^2} \int \mathrm{d}x\, \nabla\theta_\uparrow \nabla\theta_\downarrow \tag{1.91}$$

and the total Hamiltonian becomes

$$H_{\text{TLL}} = \sum_{\sigma} \int \mathrm{d}x \left[\frac{v_{\text{F}}}{2\pi} (\nabla\theta_\sigma)^2 + \left(\frac{v_{\text{F}}}{2\pi} + \frac{4V_0}{(2\pi)^2} \right) (\nabla\phi_\sigma)^2 \right] + \frac{8V_0}{(2\pi)^2} \int \mathrm{d}x\, \nabla\theta_\uparrow \nabla\theta_\downarrow. \tag{1.92}$$

The final term of Equation 1.92 breaks the quadratic form. This problem can be solved by making a unitary transformation into a new basis:

$$\begin{aligned} \theta_\rho &= \frac{1}{\sqrt{2}}(\theta_\uparrow + \theta_\downarrow) \qquad & \phi_\rho &= \frac{1}{\sqrt{2}}(\phi_\uparrow + \phi_\downarrow) \\ \theta_\sigma &= \frac{1}{\sqrt{2}}(\theta_\uparrow - \theta_\downarrow) \qquad & \phi_\sigma &= \frac{1}{\sqrt{2}}(\phi_\uparrow - \phi_\downarrow). \end{aligned} \tag{1.93}$$

In this basis, the Hamiltonian is

$$H_{\text{TLL}} = \frac{v_\rho}{2\pi} \int \mathrm{d}x \left[K_\rho (\nabla\theta_\rho)^2 + \frac{1}{K_\rho} (\nabla\phi_\rho)^2 \right] + \frac{v_\sigma}{2\pi} \int \mathrm{d}x \left[K_\sigma (\nabla\theta_\sigma)^2 + \frac{1}{K_\sigma} (\nabla\phi_\sigma)^2 \right] \tag{1.94}$$

where the various constants are

$$\begin{aligned} v_\rho &= v_{\text{F}} \sqrt{1 + \frac{4V_0}{\pi v_{\text{F}}}} \qquad & K_\rho &= \left(\sqrt{1 + \frac{4V_0}{\pi v_{\text{F}}}} \right)^{-1} \\ v_\sigma &= v_{\text{F}} \sqrt{1 - \frac{4V_0}{\pi v_{\text{F}}}} \qquad & K_\sigma &= \left(\sqrt{1 - \frac{4V_0}{\pi v_{\text{F}}}} \right)^{-1} \end{aligned} \tag{1.95}$$

Equation 1.94 is the Hamiltonian of the spinful TLL model. We see that the collective excitation of 1D fermions can be naturally separated into two bosonic modes of identical forms: in Equation 1.93 the fields of the rotated basis carry two new labels, ρ and σ. These modes can be treated as quantised density waves, which consist of the original particles of both spin orientations. However, one of them carries the net particle-number density (referred to as the charge-density wave, CDW, or holons); and the other one which carries the net spin density (referred to as the spin-density wave, SDW, or spinons). More importantly, Equation 1.95 shows that the velocities of the two modes of excitation are generally different—a remarkable result referred to as spin-charge separation, which is a hall mark of the Luttinger model.

1.3.3 Green's function of TLL

Having obtained the explicit expressions of the field operators and the Hamiltonian, we can calculate the Green's functions of the Luttinger model. As mentioned in section 1.2, a correlation function with important physical interpretation is the single-particle Green's function, which is the correlation function of single-particle creation/annihilation operators. For excitations moving in a particular direction (e. g. right movers),

$$G_{\mathrm{R}}(r) = -\Theta(\tau)\left\langle \psi_{\mathrm{R}}(r)\psi_{\mathrm{R}}^{\dagger}(0)\right\rangle . \tag{1.96}$$

In order to evaluate the time-ordered average, one needs to carry out a functional integral of the following form [90] (see also Equation 1.49):

$$\langle A(r)B(0)\rangle = \frac{1}{Z}\int \mathcal{D}\phi(x,\tau)\mathcal{D}\Pi(x,\tau)\; A(x,\tau)B(0,0) \exp\left[\int_0^{\beta} \mathrm{d}\tau \int \mathrm{d}x\,[i\Pi(x,\tau)\partial_r\phi(x,\tau) - H_{\mathrm{TLL}}]\right], \tag{1.97}$$

where Z is the partition function:

$$Z = \int \mathcal{D}\phi(x,\tau)\mathcal{D}\Pi(x,\tau)\; \exp\left[\int_0^{\beta} \mathrm{d}\tau \int \mathrm{d}x\,[i\Pi(x,\tau)\partial_r\phi(x,\tau) - H_{\mathrm{TLL}}]\right]. \tag{1.98}$$

The integral of τ over the range $[0,\beta]$ accounts for finite temperature, $\beta = 1/k_B T$ being the inverse temperature factor.

Results of the spinless model

Applying the field operator of the spinless model (Equation 1.71) to Equation 1.96 yields

$$G_{\mathrm{R}}(r) = -\frac{\mathrm{e}^{ik_{\mathrm{F}}x}}{2\pi\alpha}\left\langle \exp[i(\phi(r)-\theta(r))]\exp[-i(\phi(0)-\theta(0))]\right\rangle . \tag{1.99}$$

The evaluation of the average is complex and the details will not be produced here. The full derivation can be found in Appendix C of [27]. The result is

$$\begin{aligned} G_{\mathrm{R}}(r) &= -\frac{\mathrm{e}^{ik_{\mathrm{F}}x}}{2\pi\alpha}\exp\left[-\frac{K+K^{-1}}{2}F_1(r) - F_2(r)\right] \\ F_1(r) &= \frac{1}{2}\ln\frac{x^2+(\upsilon|\tau|+\alpha)^2}{\alpha^2} \qquad F_2(r) = -i\,\mathrm{Arg}[\upsilon\tau + \alpha\,\mathrm{Sign}(\tau)], \end{aligned} \tag{1.100}$$

where Arg[…] refers to the argument of a complex number. Two important quantities can be obtained from the Green's function. First, the occupation number (Equation 1.28) is

$$\begin{aligned} n_{\mathrm{R}}(k) &= \lim_{\tau\to 0^-} \int \mathrm{d}x\, \mathrm{e}^{-ikx} G_{\mathrm{R}}(x,\tau) \\ &= \lim_{\tau\to 0^-} -\int \mathrm{d}x\, \mathrm{e}^{i(k_{\mathrm{F}}-k)x} \frac{1}{2\pi\alpha} \exp\left[-\frac{K+K^{-1}}{2} F_1(x,\tau) - F_2(x,\tau)\right] \\ &= -\int \mathrm{d}x\, \mathrm{e}^{i(k_{\mathrm{F}}-k)x} \frac{1}{2\pi\alpha} \left(\frac{\alpha}{x^2+\alpha^2}\right)^{\frac{K+K^{-1}}{2}} \mathrm{e}^{i\,\mathrm{Arg}(-\alpha+ix)} \\ &\propto \frac{1}{2} \pm \left|k-k_{\mathrm{F}}\right|^{\frac{K+K^{-1}}{2}-1}. \end{aligned} \tag{1.101}$$

The above result, taken in the limit $\tau \to 0^-$, is the occupation number of particle-like excitations. The hole-like branch occupation number (Equation 1.29) can be obtained from the same integral with $\tau \to 0^+$. In both cases, the integral is a Fourier transform of a power law, whose functional form is $\sim \frac{1}{2} + \left|k-k_{\mathrm{F}}\right|^{\frac{K+K^{-1}}{2}-1}$ (particle) and $\sim \frac{1}{2} - \left|k-k_{\mathrm{F}}\right|^{\frac{K+K^{-1}}{2}-1}$ (hole) [83]. Figure 1.5 (left) shows a plot of the power-law distribution. Compared to the Fermi liquid distribution in Figure 1.2, the TLL is equivalent to the special case where $z_{\boldsymbol{k}} = 0$. This implies that no Fermi-like quasiparticle exists in one dimension. As the figure shows, the TLL distribution consists of equal contributions from particles and holes, making it no longer sensible to describe the system in the basis of either excitation. Instead, as was discussed in previous sections, the natural basis of the TLL are the collective modes, which are mixtures of both particles and holes.

The second quantity that can be extracted from the Green's function is the local density of states of (e. g. particle-like) excitations:

$$n_{\mathrm{R}}(\varepsilon-\varepsilon_{\mathrm{F}}) = -\frac{1}{\pi}\int \mathrm{d}t\, \mathrm{e}^{i(\varepsilon-\varepsilon_{\mathrm{F}})t}\, \mathrm{Im}\, G_{\mathrm{R}}(x=0,it) \propto \left|\varepsilon-\varepsilon_{\mathrm{F}}\right|^{\frac{K+K^{-1}}{2}-1}, \tag{1.102}$$

which again takes the form of a power law. The power law confirms again that no particle-like excitation exists in 1D, as the density of states vanishes very quickly as the excitation energy approaches the Fermi point.

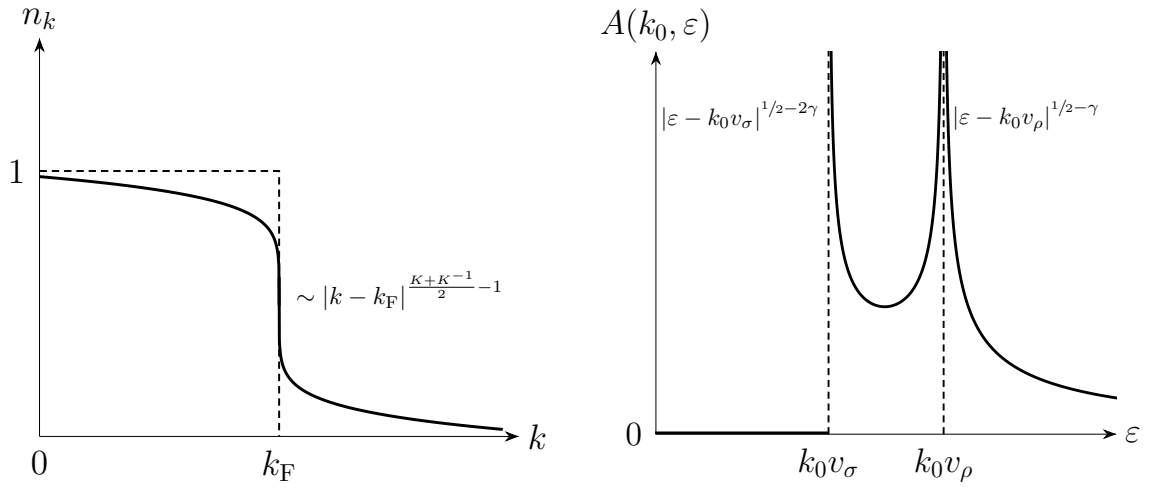


FIGURE 1.5 Left: the occupation number of hole- and particle-like excitations of spinless TLL in momentum space. The dashed lines indicate the Fermi distribution of free particles. Right: the particle branch of the spinful TLL spectral function at fixed momentum k_0. See [95, 98] for a full plot which shows variations in k.

Results of the spinful model

When the spin label is included, Equation 1.100 is modified to include equal contributions from the density and spin fields [27]:

$$G_R(r) = -\frac{e^{ik_F x}}{2\pi\alpha} \exp\left[-\frac{K_\rho + K_\rho^{-1}}{4} F_1(x, u_\rho\tau) - \frac{1}{2} F_2(x, u_\rho\tau)\right] \exp\left[-\frac{K_\sigma + K_\sigma^{-1}}{4} F_1(x, u_\sigma\tau) - \frac{1}{2} F_2(x, u_\sigma\tau)\right]. \quad (1.103)$$

The Green's function can be used to calculate $n(k)$ and $n(\varepsilon)$ as before (Chapter 7 of [27]). The remarkable result is that the local density of states, which is an easily measurable observable, does not encompass any information on the spin field:

$$n(\varepsilon) \propto |\varepsilon|^{\frac{K_\rho+K_\rho^{-1}}{4}-\frac{1}{2}} \quad (1.104)$$

This means that detecting the effects of the spin field of an TLL requires an experiment which measures a different observable. As we will see later (Equation 3.8), the energy-and-momentum-resolved tunnelling conductance measurement detects the convolution of the spectral functions between a 1D and 2D layers of electrons. We have seen in Equation 1.43 that the spectral function of electrons in a 2D layer (a Fermi liquid) is a Lorentzian function.

From the spinful Green's function, the TLL spectral function can be shown to possess the functional form [95, 63]

$$A(k,\omega) = -\frac{1}{\pi}\int \mathrm{d}x\,\mathrm{d}t\,\mathrm{e}^{-i(kx-\omega t)}\,\mathrm{Im}\,G(x,t) \propto \frac{\Theta(\omega - v_\sigma k)}{\left|\omega - v_\rho k\right|^{\frac{1}{2}-\gamma}\left|\omega - v_\sigma k\right|^{\frac{1}{2}-2\gamma}} \tag{1.105}$$
$$\gamma = \frac{K_\rho + K_\rho^{-1} - 2}{8} + \frac{K_\sigma + K_\sigma^{-1} - 2}{8},$$

which consists of two distinct power-law singularities centred at the spinon and holon energies (Figure 1.5). Again the power-law indicates that excitations in 1D differ significantly from the quasiparticles found in the Fermi liquid in higher dimensions.

1.4 Non-linear Luttinger liquid

A critical basis of the Luttinger liquid theory is that the parabolic dispersion relation of free electrons can be approximated by a linear function, as long as we restrict the discussion to low-energy excitations. The linearisation of the dispersion enables the bosonisation process, which maps fermionic excitations to bosonic fields. Electron states on a linear dispersion are equidistant, therefore fermionic excitations, which are in fact electron-hole pairs, have fixed sets of energy values $E_1, E_2, \ldots$ (Figure 1.6). They resemble a collection of bosons in some energy levels. Furthermore, each electron-hole pair has zero combined spin, similar also to a boson. Outside the low-energy limit, the finite curvature of the original dispersion means the dispersion can no longer be treated as linear. When the fermion states are no longer equidistant the bosonisation process cannot be applied and the theoretical discussion made in the last section breaks down. To differentiate them, the model in the low-energy regime is sometimes referred to as the linear Luttinger liquid, whereas any model for outside the low-energy limits is known as the non-linear Luttinger liquid.

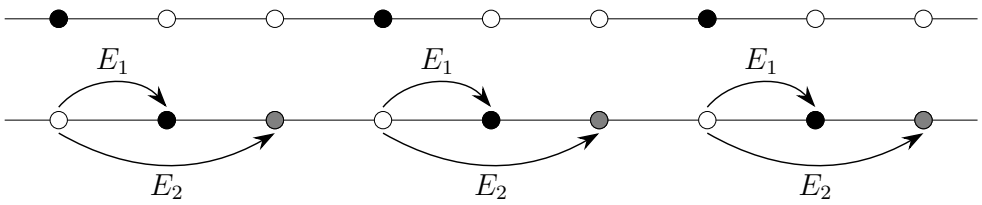


FIGURE 1.6 Top: Electron states on a linear dispersion are equidistant. The filled dots represent occupied states while the empty ones are unoccupied. Bottom: The electron-hole pairs are characterised by distinctive energies. They resemble bosons in energy levels E_1, E_2, etc.

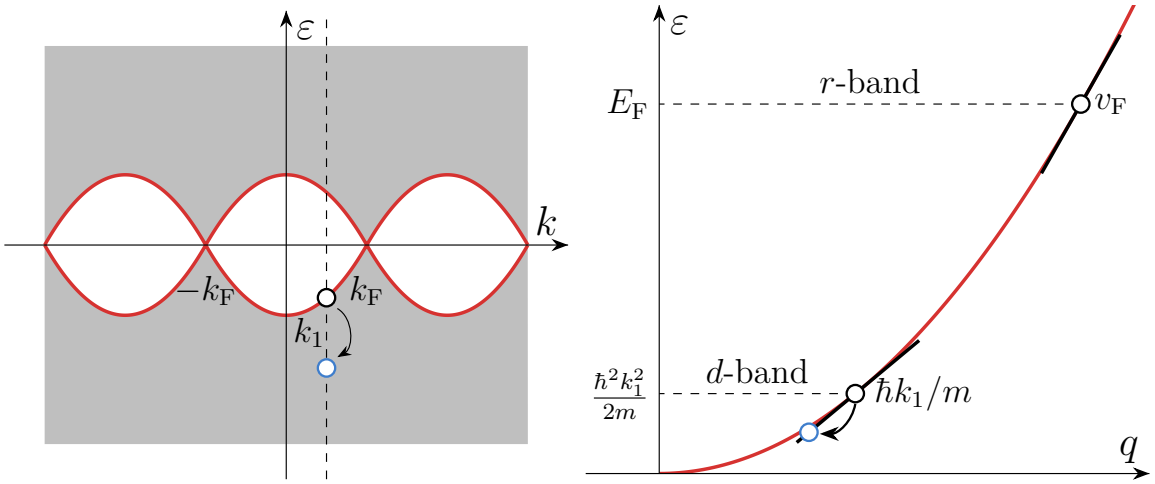


FIGURE 1.7 Left: the energy spectrum of an interacting 1D system. The grey region represents the energies accessible to the many-body excitations of the system (i. e. collections of electron-hole pairs), while the white region represents the forbidden region which arises due to the restriction of 1D geometry: a particle must move along the dispersion to be removed from a 1D chain (see also Figure 1.3 and the related text). The red curves mark the boundaries between the two regions. The white dot represents a 'deep' hole state with momentum k_1. Right: separation of the right-moving fermion excitation dispersion into two bands, one around the 'deep' hole at k_1 and the other around the Fermi energy. Images adapted from [43].

The non-linear behaviour of 1D fermions remains an open question for which a complete theoretical description is yet to be found. In the theoretical field, a technique known as the Bethe ansatz [10, 55] can be used to diagonalise many-body Hamiltonians in 1D, without making any approximations such as linearisable dispersion. However, this method involves highly complex computations, whose detailed discussion lies outside of the scope of this thesis. Additionally, the Bethe ansatz method has limited scope of applicability in the sense that its results are specific to each individual problem: the technique involves first guessing the form of the eigenfunction and proving the validity of the solution a posteriori it exists. For the problem of 1D interacting fermionic systems, Bethe anstaz computation has only been made in the case of the spinless model [91]. Nevertheless, the Bethe ansatz technique can be used to validate other models with various assumptions.

1.4.1 Mobile-impurity model

The mobile-impurity model is a phenomenological model of general applicability to a wide class of one-dimensional fermionic, bosonic, and spin systems [40]. It accounts for the non-linearity of the dispersion and describes the qualitative change in the behaviour of the spectral function. The mobile impurity model has a microscopic construction which is closely

related to the linear Luttinger model described in the earlier sections. Consider the spectrum of 1D fermionic excitations shown in the left plot of Figure 1.7. In the scenario depicted, an impurity lies far away from the Fermi energy at momentum k_1. This resembles the problem of X-ray scattering in metals, in which a hole is created deep in the conduction band when a high-energy X-ray photon is absorbed. The solution to the scattering problem is known to have a power-law singularity close to the Fermi level which arises from the interactions between the deep hole and quasiparticles near the Fermi level, calculated by the exactly diagonalisable heavy-impurity model [75, 69, 68].

The mobile-impurity model starts with the many-body fermionic Hamiltonian:

$$H = H_0 + H_{\text{int}} = \int \mathrm{d}x \left[-\frac{\hbar^2}{2m} \psi^\dagger(x) \partial_x^2 \psi(x) + V_0 \rho^2(x) \right], \tag{1.106}$$

where the interaction potential is taken to be short-range: $V(x-x') = V_0 \delta(x-x')$. It employs a construction similar to the heavy-impurity model, projecting the excitation dispersion into two bands (demonstrated by right plot of Figure 1.7), with one centred around the impurity (d-band) and the other around the Fermi level (r-band). The fermionic field operators (see also Equation 1.64) for each band are:

$$\psi_{\text{d}}(x) = \frac{1}{\sqrt{L}} \sum_{|k-q|<k_0} \mathrm{e}^{i(q-k)x} c_{q,r} \qquad \psi_r(x) = \frac{1}{\sqrt{L}} \sum_{|q|<k_0} \mathrm{e}^{iqx} c_{q,r}, \tag{1.107}$$

where $k_0 \ll k_{\text{F}}$ is the size of the bands. The subscript $r = \text{L}, \text{R}$ labels the left/right-moving fermionic excitations. Note that only right-moving excitations are involved ($r = \text{R}$), as we are only considering the right-moving branch of the dispersion. The Hamiltonian of the right-movers can be rewritten in terms of the new operators as:

$$H_{\text{R}} = \int \mathrm{d}x \Big[\underbrace{\psi_r^\dagger(-iv_{\text{F}}\partial_x)\psi_r + V_0\rho_r^2}_{r\text{-band}} + \underbrace{\psi_{\text{d}}^\dagger\Big(\frac{\hbar^2 k^2}{2m} - i\frac{k}{m}\partial_x\Big)\psi_{\text{d}} + V_0\rho_{\text{d}}\rho_r}_{d\text{-band}} \Big], \tag{1.108}$$

where the terms from each band are highlighted. The density operators are defined as $\rho_i(x) = \psi_i^\dagger(x)\psi_i(x)$. Note that since the model contains only one impurity state, there is no interaction term of the form $V_0\rho_{\text{d}}^2$ (it does not interact with itself). The prominent feature of the model is that the impurity/hole state is mobile: it follows the spectral threshold $\varepsilon_{\text{th}}(k)$, the existence of which in the many-body spectrum is a distinct property of all one-dimensional systems.

The above Hamiltonian can be solved with the same bosonisation technique that was applied to the linear Luttinger-liquid model. It can be shown [39, 73, 54] that the solution is equivalent to the following three-part phenomenological construction of the non-linear Hamiltonian, H_{nLL}:

$$\begin{gathered} H_{\mathrm{nLL}} = H_{\mathrm{TLL}} + H_{\mathrm{d}} + H_{\mathrm{int}} \\ H_{\mathrm{TLL}} = \frac{v}{2\pi}\int \mathrm{d}x \left[K(\nabla\theta)^2 + \frac{1}{K}(\nabla\phi)^2\right] \quad H_{\mathrm{d}} = \int \mathrm{d}x\, \psi_{\mathrm{d}}^{\dagger}\left[\varepsilon_{\mathrm{th}}(k) - i\frac{\partial\varepsilon(k)}{\partial k}\partial_x\right]\psi_{\mathrm{d}} \\ H_{\mathrm{int}} = \int \mathrm{d}x\, [V_{\mathrm{R}}\rho_{\mathrm{R}} + V_{\mathrm{L}}\rho_{\mathrm{L}}]\rho_{\mathrm{d}} = \int \mathrm{d}x \left(V_{\mathrm{R}}\nabla\frac{\theta-\phi}{2\pi} - V_{\mathrm{L}}\nabla\frac{\theta+\phi}{2\pi}\right)\psi_{\mathrm{d}}^{\dagger}\psi_{\mathrm{d}} \\ = \int \mathrm{d}x \left(V_\theta\nabla\theta + V_\phi\nabla\phi\right)\psi_{\mathrm{d}}^{\dagger}\psi_{\mathrm{d}}. \end{gathered} \tag{1.109}$$

H_{TLL} and H_{d} are the linear Hamiltonians of the Luttinger-liquid excitations and the impurity, respectively (see also Equation 1.85). In the interaction Hamiltonian between the two H_{int}, V_{L} and V_{R} represent the interaction constants between the impurity and the left/right-moving excitations. The validity of the second equality can be shown by substituting Equation 1.72, where θ and ϕ are the fields used in the bosonisation process. The last equality can be proved by rearranging terms, with $V_\theta = (V_{\mathrm{L}} - V_{\mathrm{R}})/2\pi$ and $V_\phi = -(V_{\mathrm{L}} + V_{\mathrm{R}})/2\pi$. The interaction constants are not free parameters. They are related to other microscopic parameters through the physical properties of the full model. One is the translational (Galilean) invariance of H_{nLL} and the other is the variation of the energy of the combined system with a small change of the particle density. The explicit relations were first derived in [39] as

$$V_\phi = \frac{1}{\pi}\frac{\partial\varepsilon_{\mathrm{th}}(k)}{\partial\rho} + \frac{v}{K} \qquad V_\theta = \frac{k}{m} - \frac{\partial\varepsilon_{\mathrm{th}}(k)}{\partial k}, \tag{1.110}$$

where m is the single-particle mass, v and K are the Luttinger parameters of the linear TLL model, and $\partial\varepsilon_{\mathrm{th}}(k)/\partial k$ and $\partial\varepsilon_{\mathrm{th}}(k)/\partial\rho$ are the derivatives of the nonlinear dispersion of the spectral edge with respect to the momentum and the density of the particles at a momentum k in the nonlinear regime.

Until now we have used the spinless Hamiltonian of the linear TLL model in Equation 1.109. The spinful version of the impurity model can be constructed in a similar manner, as shown in [76, 77]. The spinful Hamiltonian of the linear TLL contains two species of bosonic fields, one of the charge and the other is the spin degree of freedom. In non-spin-polarised systems, such as our experimental setup to be discussed later in this thesis, the impurity couples only to the charge fields since both spin components are equivalent. The non-linear

Hamiltonian was derived for the spinful case by Schmidt and Glazman in [76] as

$$H_{\mathrm{nLL}} = H_{\mathrm{TLL}} + \int \mathrm{d}x\, \psi_{\mathrm{d}}^{\dagger} \left(\varepsilon(k) - i \frac{\partial \varepsilon_{\mathrm{th}}(k)}{\partial k} \partial_x \right) \psi_{\mathrm{d}} + \int \mathrm{d}x\, (V_\theta \nabla \theta_\rho + V_\phi \nabla \phi_\rho) \psi_{\mathrm{d}}^{\dagger} \psi_{\mathrm{d}}, \quad (1.111)$$

where H_{TLL} is the spinful version of the linear Tomonaga-Luttinger model in Equation 1.94; θ_ρ and ϕ_ρ are defined in Equation 1.93. Following the same argument (as for the spinless case) on translational invariance and energy variation due to density fluctuations, the interaction constants between the impurity and the hydrodynamic modes of a Luttinger liquid are also not free parameters but are related to the other microscopic parameters of the combined model as [76]

$$V_\phi = \frac{1}{\sqrt{2}} \left(\frac{\partial \varepsilon_{\mathrm{th}}(k)}{\partial \rho} + \frac{v_s}{K_s} \right) \qquad V_\theta = \frac{1}{\sqrt{2}} \left(\frac{k}{m} - \frac{\partial \varepsilon_{\mathrm{th}}(k)}{\partial k} \right), \quad (1.112)$$

where v_s is the renormalised velocity and K_s is the dimensionless Luttinger parameter of the spin modes.

The correlation functions for the nonlinear model in Equation 1.111 can be evaluated using the path-integral formalism [40, 39]. Here we discuss only the spectral function, as it is the observable revealed by our experimental work. The explicit theoretical result obtained for the spinful model in [76] is

$$A_{\mathrm{s}}(k, \varepsilon) \sim \frac{1}{\left| \varepsilon \pm \varepsilon_{\mathrm{th}} \right|^{\alpha_\pm}}, \quad (1.113)$$

where the exponent is

$$\alpha_\pm = \frac{1 \mp 1}{2} - \frac{1}{2} \left(\sqrt{\frac{K_{\mathrm{c}}}{2}} - C_+ - C_- \right)^2 - \frac{1}{2} \left(\frac{1}{\sqrt{2K_{\mathrm{c}}}} - C_+ - C_- \right). \quad (1.114)$$

Here $\pm$ refers to the particle and hole sectors, K_ρ is the dimensionless Luttinger parameter of the charge modes, and

$$C_\pm = \frac{\dfrac{k - k_{\mathrm{F}}}{m\sqrt{K_{\mathrm{c}}}} \pm \sqrt{K_{\mathrm{c}}} \left(\dfrac{2}{\pi} \dfrac{\partial \varepsilon_{\mathrm{th}}}{\partial \rho} + \dfrac{\partial \varepsilon_{\mathrm{th}}}{\partial k} \right)}{2\sqrt{2} \left(\dfrac{\partial \varepsilon_{\mathrm{th}}(k)}{\partial k} \mp \dfrac{k_{\mathrm{F}}}{m K_{\mathrm{c}}} \right)}. \quad (1.115)$$

This exponent produced by nonlinear theory depends on momentum through a finite curvature of the spectral edge dispersion $\varepsilon_{\mathrm{th}}$, which is in stark contrast to the linear theory where the exponents are momentum-independent. The explicit form of the dispersion of the spec-

tral edge for the spinful fermions with short-range interactions was obtained in [91],

$$\varepsilon_{\mathrm{th}}(k) = \mu + \frac{k_{\mathrm{F}}^2}{2mK_s} \pm \frac{(k-k_{\mathrm{F}})^2}{2mK_s}, \tag{1.116}$$

where $\pm$ stands for the spectral edge in the particle and in the hole sectors.

In the hole sector between the $\pm k_{\mathrm{F}}$ points, where we perform the measurement in our experiment, Equations 1.114-1.116 give

$$\alpha_-(k) = 1 - \frac{K_c}{4}\left(1 - D_-(k)\right)^2 - \frac{K_c}{4}\left(\frac{1}{K_c} - D_+(k)\right), \tag{1.117}$$

where

$$D_-(k) = (k-k_{\mathrm{F}})\frac{\dfrac{k_{\mathrm{F}}}{K_c^2} + \dfrac{k}{K_s^2}}{\dfrac{k^2}{K_s^2} - \dfrac{k_{\mathrm{F}}^2}{K_c^2}} \qquad D_+(k) = (k-k_{\mathrm{F}})\frac{\dfrac{k_{\mathrm{F}}}{K_cK_s} + \dfrac{k}{K_cK_s}}{\dfrac{k^2}{K_s^2} - \dfrac{k_{\mathrm{F}}^2}{K_c}}. \tag{1.118}$$

The form of the function above depends only on two dimensionless Luttinger parameters K_s and K_c which we will use as fitting parameters. The form of the spectral function in Equation 1.113 together with the expression for the dependence of the exponent on momentum in Equation 1.117 will be used to interpret our experimental data.

Chapter 2

Literature Review

In this chapter, we review the pioneering works which led to the tunnelling spectroscopy—the experimental method central to the work presented in this thesis. Tunnelling spectroscopy involving using the empty states of a 1D/2D system to probe the dispersion and spectral function of another occupied 1D/2D system. A detailed analysis of the theories associated with this method is presented in the following chapter.

2.1 2D-2D tunnelling

The first investigation into the tunnelling between two 2D systems was reported by Somliner *et al.* in [85], which measured the tunnelling of electrons of an accumulation layer and a Schottky barrier into the 2DEG of a GaAs/AlGaAs high electron mobility transistor (HEMT) structure and studied the momentum selection rules of the tunnelling. Since then there have been increasing number of works on 2D-2D tunnellings in heterostructure samples. The double 2DEG scheme, which forms the experimental basis of this work, saw early application in the works of Somliner and Demmerle [84, 19], where conductance was measured on a sample with two parallel 2DEGS separated by a $\sim$ 20 nm barrier. Each 2DEG connected to separate source/drain contacts, thereby creating a current path which forces tunnelling between the two quantum wells. The independent connection to the Ohmic contacts was achieved by a combination of a lightly diffused Ohmic contact and selective etching of the sample material from the top surface (left panel, Figure 2.1). Later on, Eisenstein *et al.* [24, 23] used Schottky gates on the top and back surfaces on a double 2DEG sample that is chemically thinned to 55 μm to achieve independent contacts, as well as to control the carrier densities of the double 2DEGs, thereby providing a means to adjust the Fermi vectors of the two layers (right panel, Figure 2.1). These works investigated tunnelling in the equilibrium tunnelling regime, where the Fermi energies of the double layers are close to each other. They showed that res-

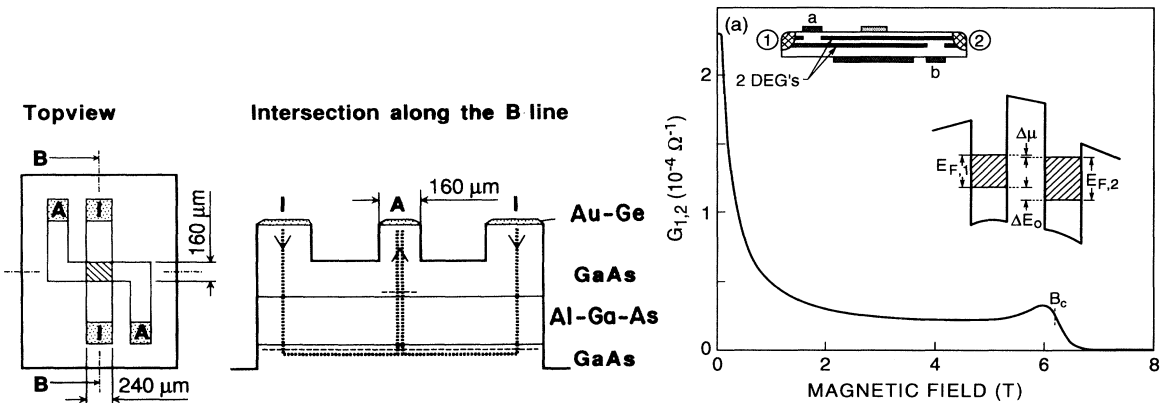


FIGURE 2.1 Left: Schematic of the 2D-2D tunnelling device reported in [19]. The top/bottom 2DEG layers are contacted separately by the Ohmics labelled A and I, respectively. 2D-2D tunnelling occurs in the overlap area, highlighted by dashed lines in the top-view; the dotted line in the cross-sectional-view indicates the current path. Right: Single magnetic field sweep of 2D-2D tunnelling conductance measured at T =0.3 K, with slightly different densities in the 2DEG layers, as reported in [23]. Cross sectional device schematic is shown in the top inset. Conductance band structure is shown in the right inset, illustrating that tunnelling occurs when the Fermi surfaces in the two layers are aligned. $\Delta\mu$ represents the excitation voltage used for measurement.

onant tunnelling is the result of transitions between quantised states in the double 2DEGs. Additionally, application of parallel magnetic field showed the conservation of canonical momentum to be a crucial part of the tunnelling condition, while perpendicular field showed tunnelling between Landau levels.

Advancement in MBE equipment and techniques brought improvements to the double 2DEG design. Brown reported [15] a device featuring back gates fabricated *in situ* using focused ion beam, where Ga ions were implanted into an MBE wafer with a 65 nm layer of n^+ GaAs, resulting in insulating regions. The layer is patterned and serves as the backgate. The double 2DEG structure is grown directly top of the patterned GaAs layer by MBE after the ion beam process. This method eliminated the need for difficult fabrication steps such as sample thinning and shallow contacts and has higher yield than the previous method. Following the work by Eisenstein into the strong magnetic field limit [25], Brown used the newer device to investigate 2D tunnelling up to 16 T [16] and later 25 T [14]. In the strong field limit all the electrons in a 2DEG are forced into the lowest Landau level. These studies observed strong suppression of the tunnelling peak at zero-bias, and the appearance of very broad tunnelling peaks at higher biases. At low filling factors, the suppression appears as a gap in the tunnelling current-voltage plot known as the Coulomb gap. This is interpreted as an effect due to electron-electron interactions in the 2DEG. A schematic of the device and results are shown in Figure 2.2.

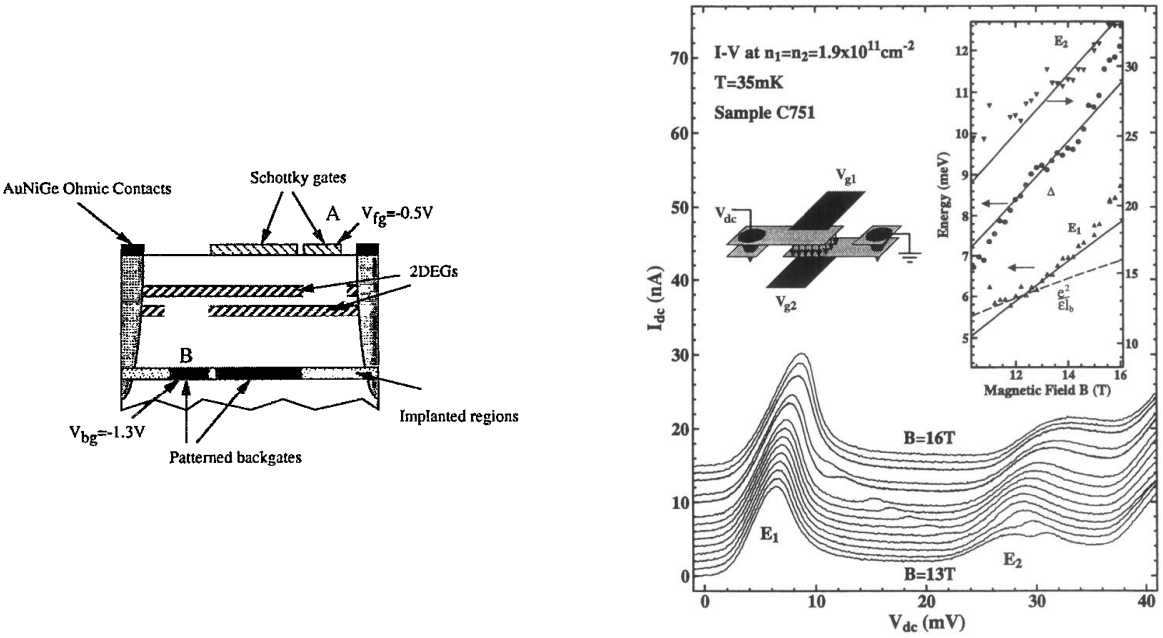


FIGURE 2.2 Left: Cross-sectional schematic of the tunnelling device with ion-beam defined backgates reported in [15]. Right: I-V characteristic of 2D-2D tunnelling current at high field reported in [16], showing suppression of conductance (Coulomb gap) at $V_{dc} = 0$ mV, a broad peak (E_1) at ~8 mV and another (E_2) at ~29 mV. The inset shows the variation of the peak positions and the gap size in terms of magnetic field.

A detailed review on 2D-2D tunnelling was given by Turner [94], combining the data from the high field measurement of Brown with additional results at intermediate fields. The tunnelling current was calculated using Bardeen's tunnelling Hamiltonian formalism, which will be discussed in the next chapter. The main point of the formalism is that the tunnelling current is proportional to the convolution of the spectral functions of the two layers involved in the process. The Lorentzian spectral function of a function of a 2D system implies the current also has a Lorentzian lineshape. Data from both equilibrium and non-equilibrium tunnelling showed that the tunnelling conductance is best fitted by a Lorentzian function (in comparison to a Gaussian and the derivative to a Fermi function), indicating that the spectral functions of both 2D layers involved in the tunnelling are Lorentzians. The line-width of the Lorentzian at half maximum, Γ, corresponds to the sum of the width of the spectral functions of the two layers. It is related to the average scattering rate of electrons of the two-layer system, τ, which satisfies the relation $\tau = \hbar/\Gamma$. The overall scattering rate has contributions from electron-impurity, electron-phonon, and electron-electron scatterings. It was shown that Γ has temperature dependence that is mainly due to electron-electron scattering and is proportional to T^2; at low temperature the electron-impurity scattering dominates, which has a dependency on the carrier density and is proportional to $n^{1/2}$. It was noted that the

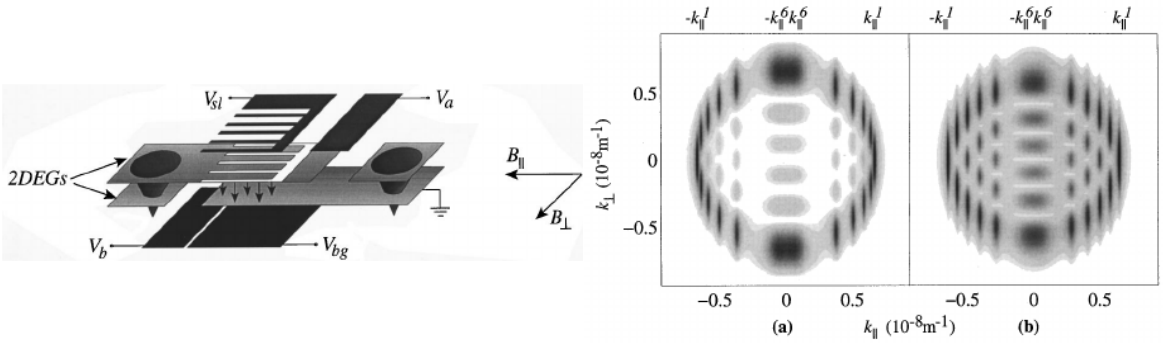


FIGURE 2.3 Left: Schematic of 1D-2D magneto tunnel device [52]. Right: 1D spectral function reported in [51], calculated by assuming a parabolic confinement potential (left) and a square well potential (right).

capacitance effects must be considered in the formalism: At finite source-drain bias voltage there will be a density difference between the two layers due to capacitive effect, which leads to a shift in the Fermi energies of the layers. This results in an enhanced linewidth at finite source-drain bias compared to equilibrium, which was reported to be typically an 18% increase. In the presence of a perpendicular magnetic field, oscillations which resemble the Shubnikov de Haas effect was observed in the tunnelling conductance, which can be attributed to the tunnelling between Landau levels in the 2D density of states. A suppression of tunnelling conductance can be observed starting from $B = 0.6\,\mathrm{T}$. As the filling factor ν approaches one half at high field, the previously mentioned Coulomb gap appears. It was shown that this gap has a linear dependency in the magnetic field when $\nu = 1/2$, which becomes a sublinear power at higher half-integer filling factors.

2.2 1D-2D tunnelling

Following the works on 2D-2D tunnelling, Kardynal *et al.* used double quantum well spectroscopy to study the tunnelling between an 1D surface superlattice and a 2DEG. Their device featured ion-beam-fabricated backgates, and contained two 18 nm quantum wells separated by a 12.5 nm barrier [52] (left panel Figure 2.3). The surface features a comb-shaped array of long and narrow gates, used to induce a 1D superlattice confinement potential in the upper 2DEG. As the upper layer is modulated, the lower 2DEG is unperturbed due to screening. The superlattice consists of 50 repeat units that are 30 μm long. Two lattice periods were studied: 350 nm (180 nm wire width) and 570 nm (340 nm wire width), and the results were presented in [50, 51]. By comparing the experimental data against the conductance calculated using the transfer Hamiltonian formalism, it was shown that the tunnelling conductance can be used to reveal structures of the 1D spectral function. Specifically, Kardynal demonstrated that the

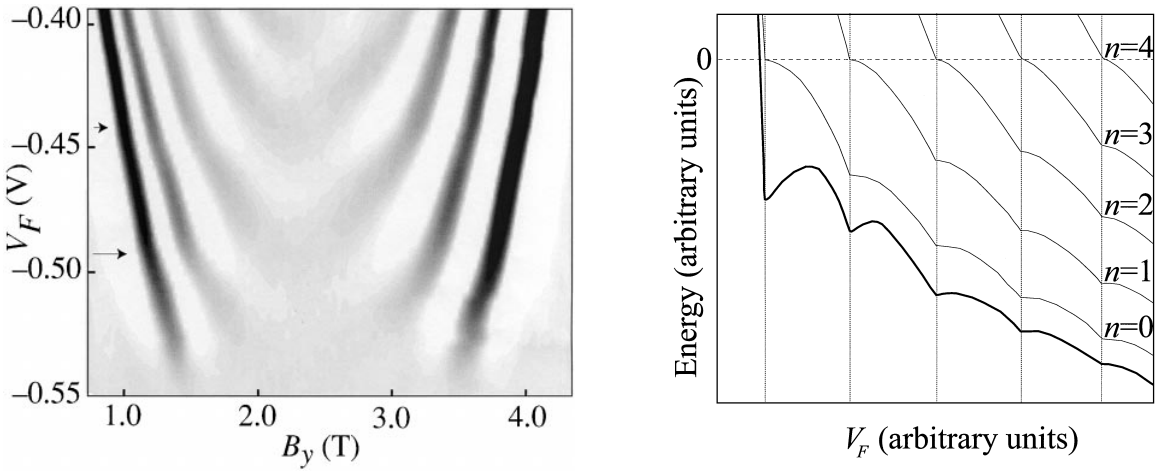


Figure 2.4 Left: Grey scale plot of the tunnelling conductance under sweep of perpendicular magnetic field. The arrows highlight front-gate voltages (V_F) where the fourth and the third 1D subbands depopulate. Right: Minimum of the confinement potential (bottom bold curve), and the energies of the first five 1D subbands calculated using the Hartree model. Dashed horizontal line represents the Fermi level. Figures taken from [60].

1D spectral function, when calculated under the assumption of a square confinement potential in the superlattice, is noticeably different than that of a parabolic potential (right panel Figure 2.3). Using in-plane magnetic field sweep both parallel and perpendicular to the 1D wires, the tunnelling conductance can be used to map out, with a high degree of sensitivity, the 1D spectral function, and consequently deduce the confinement potential lineshape up to the Fermi level. The study showed that the confinement in narrow channels is essentially parabolic, while in wide channels it is a smoothly varying function which resembles a square function. Due to the sensitivity of the conductance to the spectral function, the study also proposed that the device be used for investigation into the non-Fermi-liquid behaviour of 1D systems, which eventually led to the investigation into TLL by Jompol [44] and the work described in this thesis.

Using the same 1D-2D tunnelling device design as Kardynal, Macks *et al.* studied an effect that is observed during depletion of the 1D subbands [60]: when the top surface gate is made increasingly negative, bumps can be seen in the conductance peaks at voltages that coincides with the depopulation of 1D subbands. This is shown in the left panel of Figure 2.4, where the bumps are highlighted by arrows. Calculation using the Hartree model (right panel Figure 2.4) shows the variation of the subband energies and the confinement potential in response to the front gate voltage. As shown by the figure, when the front gate voltage (V_F) is made increasingly negative, whenever a subband is depopulated, it is pinned at the chemical potential. This leads to the oscillations of the Fermi energy, which is equal to the negative of the

confinement potential, represented by the bold curve at the bottom of the figure. This study shows that the Fermi energy of a 1D system does not vary smoothly nor monotonically with the gate voltage in surface depletion techniques.

The feasibility of using 1D-2D tunnelling spectroscopy to detect Luttinger liquid behaviour was studied by Altland in [1]. Assuming the drag effects between the 1D and 2D layers are negligible (valid for low temperature) and that tunnelling occurs between neighbouring points, the paper showed that the leading order of tunnelling conductance is given by Equation 3.3. Substituting a 2D spectral functions with no negligible disorder and interactions and a 1D spectral function given by [95], it was shown that the I-V characteristics of the tunnelling current is essentially determined by the two spin-charge parameters of the Luttinger liquid model. Additionally, it was pointed out that the smearing effect due to temperature and disorder must be much smaller than any characteristic energies in order for the structures of the I-V characteristic to be resolved.

2.3 1D-1D tunnelling

Up to now we have described works where a 2DEG is used as the probe to detect the spectral features of 1D wires. The advantage of this scheme is that the 2DEG has a simple and well-understood model: its many-body excitation is single-particle like. This reduces the complexity of any modelling work on the tunnelling. However, it is also possible to probe 1D physics using another 1D wire as the spectrometer. One such experiment was performed by Auslaender in [4, 3]. The sample was fabricated using cleaved edge over growth [71], a technique for making high quality quantum wires (with mean free path over 10 μm). As shown in Figure 2.5, two parallel 1D wires are formed at the cleaved edge of a double quantum well sample, separated by a 6 nm inter-well barrier. Only the upper well is populated by modulation doping. Quantum wires are formed by triangular confinement potential at the edge of the wells following cleaved edge overgrowth, and runs along the whole side of the sample. Two surface gates are used to deplete and dived the 2DEG in the upper well into two sections. They are coupled to the upper wire and are used to as source and drain contacts. Tunnelling occurs between the quantum wires. Because the tunnel junction defined between the gates (length $\boldsymbol{L}$ in the figure) is much shorter than the other junction, the conductance measured between the source drain contacts is dominated by the conductance from the short junction. In the conductance plot reported in [4] (top left), the calculated non-interacting dispersions do not fit the data very well. For a better match, a different effective mass (found to be $0.75m*$) must be used, and the authors interpreted that being consistent with a Luttinger liquid behaviour with interaction parameter $g = 0.75$. A suppression of conductance

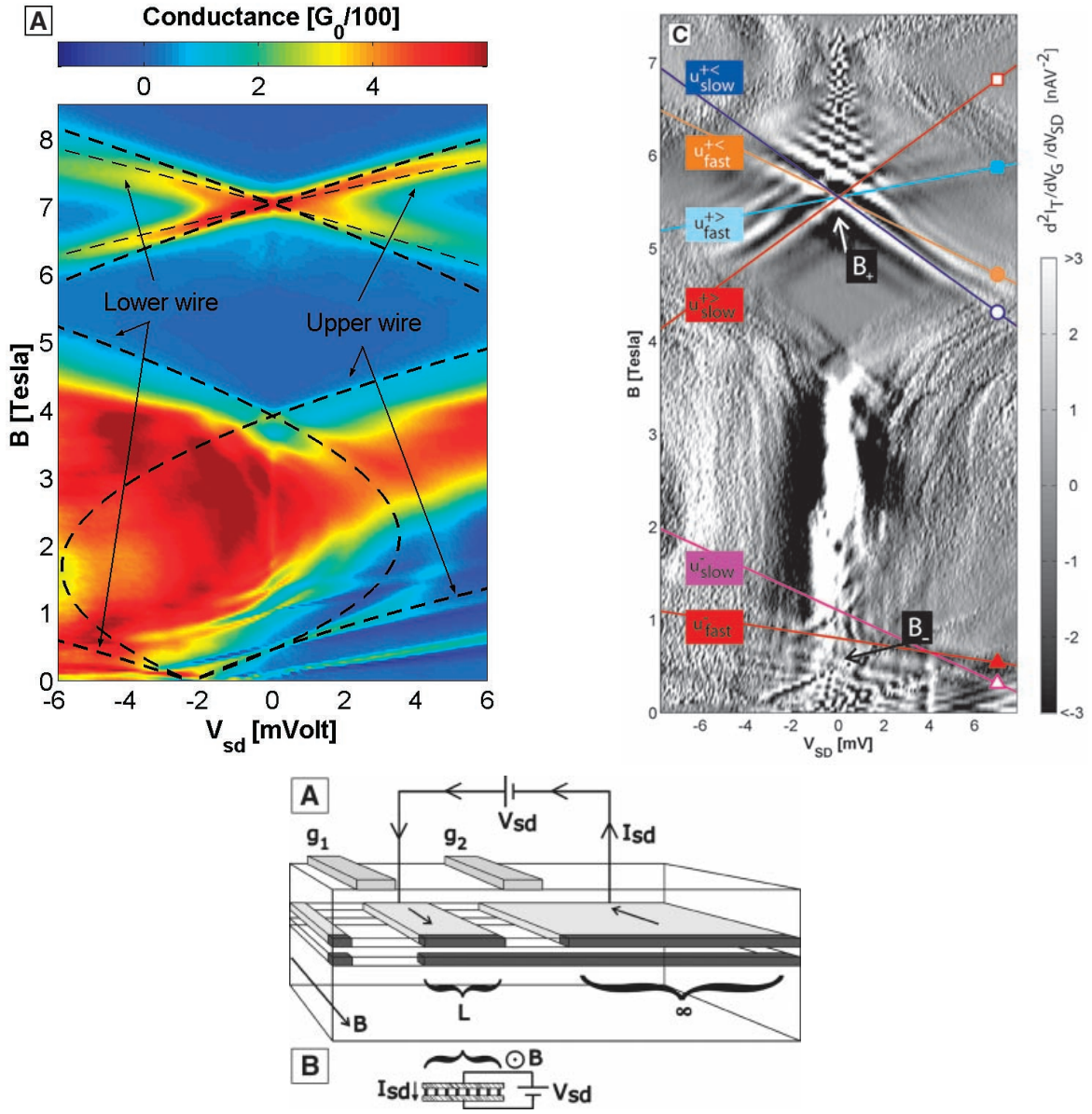


Figure 2.5 Upper left: Density plot of the tunnelling conductance scan reported in [4]. Bold dashed curves are dispersions calculated with non-interacting theory. Thin dashed curves are dispersions calculated with interacting theory. Upper right: Density plot of numerical derivative with respect to B of the tunnelling conductance reported in [3]. Taken at gate voltage −3 V, where only the last 1D subband is occupied. Lower: Schematic of the cleaved edge overgrowth device reported in [4].

near zero source-drain bias (known as the zero-bias anomaly) was also observed and is also considered a hallmark of Luttinger liquid.

In the experiment of [3], the device had an additional 2 µm wide surface gate over the 2DEG to control its density. An AC voltage was added to modulate the control voltage of this surface gate, so that the measurement is sensitive only to the part of the tunnelling junc-

tion that is affected by the control voltage. The authors observed 6 sets of slopes near the zero bias line, each corresponding to an apparent velocity of a given mode. A set of scans was obtained while changing the top gate voltage. Upon subtraction of capacitance effect, it was found that within the range where only the last 1D subband was occupied, all the fast branch velocities fall on the same curve in terms of density, while the slow branch velocities fall onto two curves. Following the model in [62], it was shown that there are two charge mode velocities, one for symmetric excitations and one for anti-symmetric excitations, with the symmetric velocity being higher than the anti-symmetric counterpart. Data from the experiment showed that all the capacitance corrected fast velocities corresponds to the calculated anti-symmetric velocity, while curiously the symmetric velocity is completely absent; The slow velocities, which corresponds to the spin mode excitations, were found to be lower than the Fermi velocity, which suggests strong back scattering.

2.4 Experimental studies on Luttinger liquid behaviour

Most experiments that have been carried out to probe for TLL behaviour focus on two important properties of the TLL: the first one is that all correlation functions behave as power laws [28], which depend on two parameters—the velocity of the density wave, v_c, and the Luttinger parameter, K. While the exponents of the correlation functions are non-universal, they are all functions of the parameter K, meaning there is some universal relationship between the exponents. This implies that one can always extract the K value from some measurement, which serves as a good description of the system. The second property is the separation of the good excitations into holons and spinons (which carry, separately, only charge and spin). In earlier texts 1.3.1 the formulations of the bosonisation method have been discussed. This separation of excitations is in fact a general consequence of the reduced freedom of motion in 1D called fractionalisation: in higher dimensions, excitations that are created by changes in some quantum number have a minimum possible value which is generally stable. In one dimension, however, excitations can generally be viewed as composites of smaller quantum numbers. An example is a magnetic chain which has anti-ferromagnetic ordering (Figure 2.6). In higher dimensions, an excitation can be created locally by flipping the spin at a particular site, creating a magnon with magnitude $\Delta S = 1$. In 1D, however, it is impossible to flip a spin without disturbing the neighbouring site, hence it is not possible to create a magnon locally. Instead, a magnon is composed of two collective excitations called spinons, each with half the magnitude of the magnon, which is created by flipping all the spins to the right of a position.

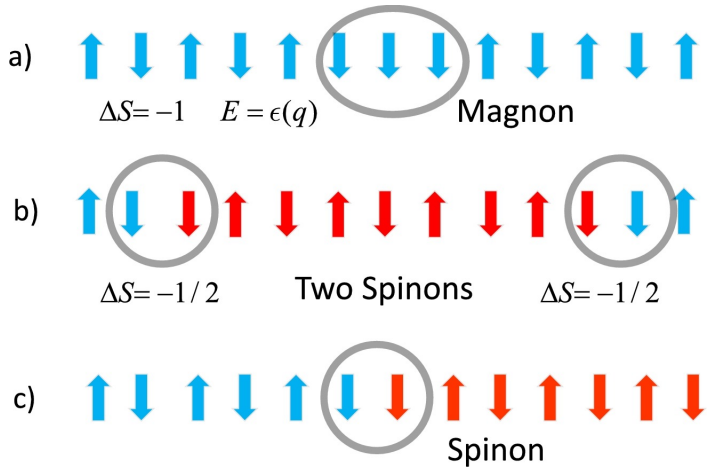


FIGURE 2.6 Magnetic chain which demonstrates fractionalisation. a) is the higher dimension case, where a magnon can be created locally by flipping the spin at a certain site. b) is the 1D case, where a magon cannot be created locally, but instead must be thought of as the sum of two composite spinons of half spin. c) is a single spinon, which is created by flipping all the spins to the right of a particular site. Image from [29].

An early experiment to observe power-law behaviour of a TLL comes from the measurement of high-frequency optical conductivity of several organic conductors/superconductors [80, 42, 57], which are anisotropic materials with greatly different hopping potentials along three perpendicular directions. The one-dimensional physics is probed with optical frequency, which can easily exceed the effective coupling towards the first transverse direction, but cannot be achieved with temperature variation. Here the optical conductivity exhibits power-law dependence on the frequency. Shortly afterwards, temperature and bias-voltage power laws were observed in tunnelling conductance measurements of carbon nanotubes, which were in good agreement with the expected behaviour of a TLL [13]. This experiment also showed variations in the power law, depending on the whether the tunnelling occurred at the end or in the middle of the tube. Tunnelling inside a TLL can also be studied in bulk material purple bronze compound ($Li_{0.9}Mo_6O_{17}$) by photo-emission spectroscopy [20, 35, 99, 33], where power-law behaviour is observed in the density of states. Another system where TLL physics may be applicable is the edge states that are responsible for the fractional quantum-Hall effect [18]. The two chiral states that exist at the edge of a Hall bar can be treated as the left and right movers of the TLL model, and the TLL parameter K can be shown to be related to the fractional quantum Hall plateau [27]. An artificial system of cold atoms is another class of material where a TLL may be achieved [12]. These systems work by trapping neutral atoms in an artificial lattice of interfering lasers. By controlling the power of the lasers, one is able to adjust the hopping potentials in any chosen directions, which allows one-dimensional systems to be created in a manner similar to the organic conductors discussed earlier.

In terms of the detection of fractionalisation, an early experiment was the measurement of the excitation spectrum in a quantum spin chain system (antiferromagnetic material $KCuF_3$) by neutron scattering, where the fractionalisation of magnons into two spinons is detected. Beyond spin chain systems, the best observation of the spin charge separation was made in tunnelling experiments, for example, [4], where electrons are directed to tunnel between the Fermi points from one quantum wire to another, under the control of gate voltage and magnetic field; and [44], where electrons tunnel between the Fermi points of one quantum wire and a quantum well. In both experiments, two excitation modes with distinctive velocities were observed.

Chapter 3

Theory of energy and momentum-resolved tunnelling

In this chapter, we discuss the theory of energy and momentum-resolved magneto-resonance tunnelling in a heterostructure, central to the experimental work of this work. Similar to the works [50, 52, 51] reviewed in the previous chapter, our experiment involves a heterostructure sample with two closely-spaced parallel quantum wells, whose upper layer is modulated by an array of 1D confinement potential from surface gates while the lower layer remains unperturbed. The current of electrons that tunnel between the two quantum wells can be used to determine the dispersion of the electrons in both systems, as well as reveal information about their spectral functions.

3.1 Transfer Hamiltonian formalism

The transfer Hamiltonian formalism is a method for calculating the current of quantum tunnelling, originally developed by Bardeen for a study of metal-insulator-metal junctions [8]. The current is taken to be proportional to the tunnelling rate and the local densities of the states before and after the tunnelling event. Assuming electrons tunnel from the upper layer (subscript 1) into the lower layer (subscript 2), the rate of tunnelling (i. e. tunnelling probability) is given by Fermi's golden rule (see also Equation 1.5):

$$w = \frac{2\pi}{\hbar}\left|T_{\boldsymbol{k}_1\to\boldsymbol{k}_2}\right|^2 \delta(E_1 - E_2). \tag{3.1}$$

A full derivation of the expression of tunnelling current can be found in the review [59]. We remark that the delta function term enforces the conservation of energy: tunnelling occurs only when the initial and final states have the same energy. $T_{\boldsymbol{k}_1\to\boldsymbol{k}_2}$ is known as the transmis-

sion matrix, which is given by $T_{\boldsymbol{k}_1 \to \boldsymbol{k}_2} = \langle \psi_1 | U | \psi_2 \rangle$, where U is the potential function of the tunnel barrier. Following Bardeen's calculation, the matrix element turns out to be given by the integral:

$$T_{\boldsymbol{k}_1 \to \boldsymbol{k}_2} = \frac{\hbar^2}{2m^*} \int \mathrm{d}x\, \mathrm{d}y \left(\psi_2 \frac{\partial \psi_1^*}{\partial z} - \psi_1^* \frac{\partial \psi_2}{\partial z} \right). \tag{3.2}$$

We define the x-y plane to be that of the sample surface, with the x axis parallel to the 1D channels. Combining the tunnelling rate and the density of states, we have the expression for the tunnelling current [79, 1]:

$$I = \frac{4\pi e}{\hbar} \sum_{\boldsymbol{k}_1} \sum_{\boldsymbol{k}_2} \left| T_{\boldsymbol{k}_1 \to \boldsymbol{k}_2} \right|^2 \int \mathrm{d}E\, A_1(\boldsymbol{k}_1, E) A_2(\boldsymbol{k}_2, E + eV_{\mathrm{DC}}) \left[f(\varepsilon_1) - f(\varepsilon_2 + eV_{\mathrm{DC}}) \right], \tag{3.3}$$

where $f(\varepsilon_i)$ is the Fermi-Dirac distribution, and $\varepsilon_i = E - \mu_i$ is the electron energy (E) relative to the Fermi level (μ_i). The subscript $i = 1, 2$ labels the upper and lower layers respectively. The densities of states of the two layers are given by the spectral functions A_1 and A_2 (see Equation 1.102 and Equation 1.105). The last term of the integral is the difference between two Fermi-Dirac distributions offset by eV_{DC}, where V_{DC} is a DC bias voltage applied across the two layers, used to control the offset between their Fermi energies. At very low temperatures, the Fermi-Dirac distribution is approximately a step function. For a function $g(E)$ that varies slowly with the temperature (i. e. almost constant), one can make the approximation $\int \mathrm{d}E\, g(E)[f(\varepsilon_1) - f(\varepsilon_2 + eV_{\mathrm{DC}})] \approx g(E) eV_{\mathrm{DC}}$ (see Figure 3.1). Applying this result, Equation 3.3 simplifies to

$$I = \frac{4\pi e^2 V}{\hbar} \sum_{\boldsymbol{k}_1} \sum_{\boldsymbol{k}_2} \left| T_{\boldsymbol{k}_1 \to \boldsymbol{k}_2} \right|^2 A_1(\boldsymbol{k}_1, \varepsilon_1) A_2(\boldsymbol{k}_2, \varepsilon_2 + eV). \tag{3.4}$$

This equation suggests that the tunnelling current is proportional to the convolution between the spectral functions of the two layers.

We can evaluate the tunnelling matrix element with the electron wavefunctions of the two layers:

$$\psi_1(x, y, z) = \mathrm{e}^{ik_x' x} \phi_n(y) \chi_1(z) \qquad \psi_2(x, y, z) = \mathrm{e}^{ik_x x} \mathrm{e}^{ik_y y} \chi_2(z). \tag{3.5}$$

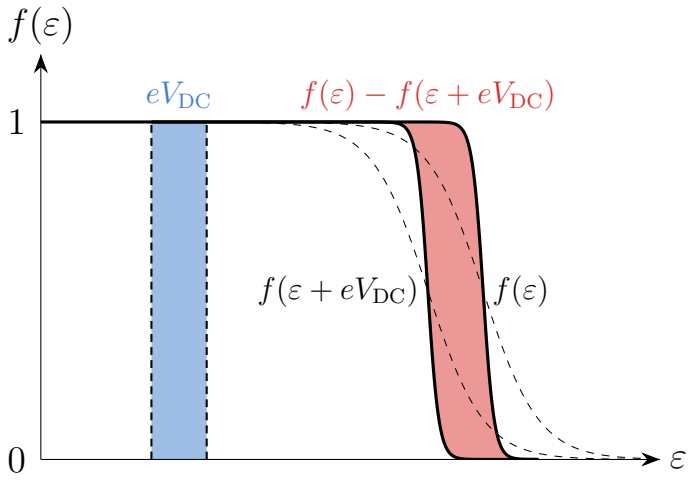


FIGURE 3.1 At low temperatures, the difference between two Fermi-Dirac distributions offset by a DC bias V_{DC} (red area) is approximately equal to the energy offset eV_{DC} multiplied by 1 (blue area). When $T = 0$, the two areas become identical. Conversely, the approximation is worse at higher temperatures (e. g. compare the area enclosed by the dashed profiles versus the blue rectangle).

Substituting these into Equation 3.2 yields

$$T_{\boldsymbol{k}_1 \to \boldsymbol{k}_2} = \frac{\hbar^2}{2m^*}\left(\chi_2 \frac{\partial \chi_1^*}{\partial z} - \chi_1^* \frac{\partial \chi_2}{\partial z}\right) \int \mathrm{d}x\, \mathrm{e}^{i(k_x - k_x')x} \int \mathrm{d}y\, \phi_n(y) \mathrm{e}^{ik_y y}$$
$$= (2\pi)^2 t(z) \delta(k_x - k_x') \tilde{\phi}_n(k_y), \quad (3.6)$$

where $t(z) = \frac{\hbar^2}{2m^*}\left(\chi_2 \frac{\partial \chi_1^*}{\partial z} - \chi_1^* \frac{\partial \chi_2}{\partial z}\right)$. The delta function and $\tilde{\phi}(k_y)$ result from the x and y integrals, respectively, which are both Fourier transforms. The delta function suggests that, just like the energy, momentum in the x direction is conserved in the tunnelling process. We treat the confinement potential induced by the gates which define the quantum wires in the upper layer as parabolic: $V(y) = \frac{1}{2} m^* \omega_0^2 y^2$, where y is the spatial coordinate. In this case, the wavefunction of the n-th eigenstate in the y direction is that of a quantum harmonic oscillator [11], also known as the Hermite function:

$$\phi_n(y) = \frac{1}{\sqrt{2^n n! l \sqrt{\pi}}} H_n\left(\frac{y}{l}\right) \exp\left(-\frac{y^2}{2l^2}\right), \quad (3.7)$$

where $l = \sqrt{\hbar / m^* \omega_0}$ is a scaling factor known as the natural length, which makes the exponent dimensionless, and H_n represents the n-th degree Hermite polynomials. A well-known result of the harmonic oscillator is that the energy spacing between adjacent subbands is $\hbar\omega_0$, which will be used later to determine the value of ω_0 experimentally. The tunnelling matrix is

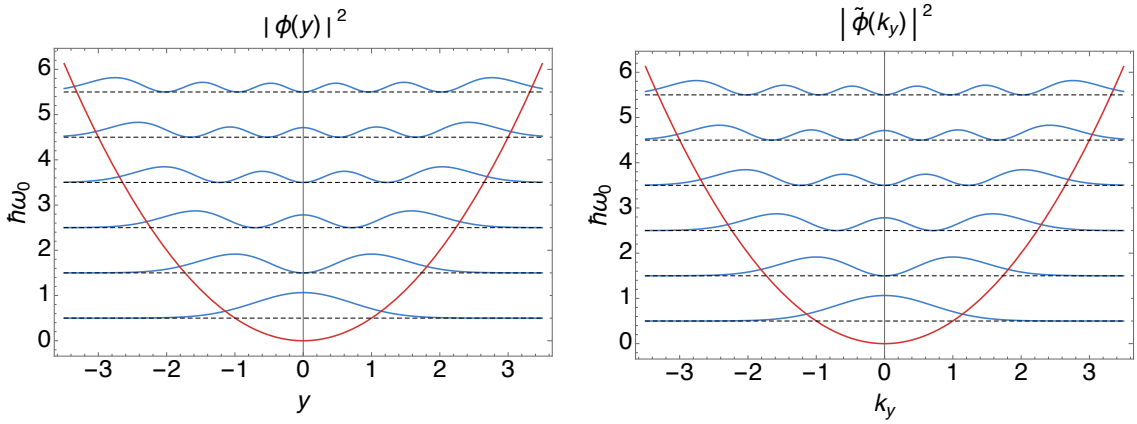


FIGURE 3.2 The probability distribution in position $\left|\phi_n(y)\right|^2$ and in momentum $\left|\tilde{\phi}_n(k_y)\right|^2$ of the n-th eigenstates in a parabolic potential. The left plot is produced with parameters $m^*\omega^2 = 1$ and $\hbar/m^*\omega_0 = 1$. The Fourier transform is calculated by Mathematica. Both distributions are normalised explicitly by their integrand over the whole argument domain.

proportional to $\tilde{\phi}(k_y)$, which is the Fourier transform of the wavefunction $\phi_n(y)$. Physically it represents the wavefunction in momentum space. We plotted in Figure 3.2 the probability distributions $|\phi(y)|^2$ and $\left|\tilde{\phi}(k_y)\right|^2$, which shows the remarkable result that the two functions have identical line shapes. Mathematically, this is because the whole class of Hermite functions (defined as the product between a Hermite polynomial and a Gaussian function) forms an orthonormal eigen-basis to the Fourier transform operator [56], which means they are unaltered by Fourier transforms.

We have seen that a necessary condition for tunnelling (also called the resonance condition) is that the electron states before and after tunnelling must be matched in energy and x-momentum. While the energy offset between the two layers can be controlled by a DC bias, the x-momentum offset can be varied by placing the sample in a magnetic field aligned to the y-direction. In the presence of a magnetic field, the canonical momentum of an electron is $\boldsymbol{p} = \hbar\boldsymbol{k} + e\boldsymbol{A}$, where $\boldsymbol{A}$ is the magnetic vector potential that satisfies $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. One can verify that a magnetic field $\boldsymbol{B} = (0, B, 0)$ is consistent with the choice of the Landau gauge $\boldsymbol{A} = (zB, 0, 0)$. In this gauge the momentum boost caused by the magnetic field is $\hbar\Delta\boldsymbol{k} = e\boldsymbol{A} = (eBd, 0, 0)$, where d is the inter-layer distance in the z-direction, which is nominally 32 nm for our wafers.

In summary, we expect the magneto-resonance tunnelling current to be

$$I \propto \int \mathrm{d}E\,\mathrm{d}\boldsymbol{k}\,[f(\varepsilon_1) - f(\varepsilon_2 + eV)]\left|\tilde{\phi}_n(k_y)\right|^2 A_1(\boldsymbol{k}, E) A_2\left(\boldsymbol{k} + \frac{ed}{\hbar}\tilde{\boldsymbol{n}} \times \boldsymbol{B}, E + eV\right), \quad (3.8)$$

where $\tilde{\boldsymbol{n}}$ is the unit vector in the direction of the current. We have replaced the DC bias term with eV, where $V = V_{\mathrm{DC}} + V_{\mathrm{AC}}$, to reflect the experimental setup (Section 6.1.2) in which the sample is measured with a lock-in amplifier, which applies an additional AC signal across the sample. The summation in k-space has been replaced by integrals, which is valid due to the large number of states. We stop keeping track of the prefactors from now on, as we will only be interested in the functional form of the current.

3.1.1 Evaluating the tunnelling conductance

Due to phase sensitive measurement, the quantity detected by the experiment is in fact the differential conductance of the sample. We therefore need to derive the expression for the conductance from Equation 3.8. Let f represent the energy integral of the spectral functions:

$$f = \int_{E_1}^{E_2} \mathrm{d}E\, A_1(\cdots, E) A_2(\cdots, E + eV). \tag{3.9}$$

According to Equation 3.8, $I \propto \int f\, \mathrm{d}\boldsymbol{k}$. The differential conductance is therefore

$$G = \frac{\mathrm{d}I}{\mathrm{d}V} \propto \frac{\mathrm{d}}{\mathrm{d}V} \int f\, \mathrm{d}\boldsymbol{k}\,. \tag{3.10}$$

Consider the following definite integral,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} \int_{g(\alpha)}^{h(\alpha)} \mathrm{d}x\, f'(x, \alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha} \left[f(x, \alpha) \Big|_{x=g(\alpha)}^{x=h(\alpha)} \right] = \left[\frac{\partial f(x, \alpha)}{\partial \alpha} + \frac{\partial f(x, \alpha)}{\partial x} \frac{\partial x}{\partial \alpha} \right] \Bigg|_{x=g(\alpha)}^{x=h(\alpha)}, \tag{3.11}$$

where f' denotes $\partial f \big/ \partial x$. The derivatives in the first two expressions are total derivatives because there is no x dependence in the functions to be differentiated (the x integral is fixed by boundary values). In the context of our conductance derivation, we can make the substitutions $x = E$, $\alpha = V$ and $f'(x, \alpha) = A_1(\cdots, E) A_2(\cdots, E + eV)$. Then Equation 3.11 shows that the derivative $\mathrm{d}f \big/ \mathrm{d}V$ in Equation 3.10 evaluates to

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}V} \int_{E_1}^{E_2} \mathrm{d}E\, A_1(\cdots, E) A_2(\cdots, E + eV) &= \frac{\partial}{\partial V} \int_{E_1}^{E_2} \mathrm{d}E\, A_1(\cdots, E) A_2(\cdots, E + eV) \\ &+ A_1(\cdots, E_2) A_2(\cdots, E_2 + eV) \frac{\partial E_2}{\partial V} - A_1(\cdots, E_1) A_2(\cdots, E_1 + eV) \frac{\partial E_1}{\partial V}. \end{aligned} \tag{3.12}$$

This result can be reorganised by introducing two functions:

$$P_1 = \frac{\partial}{\partial V}\int_{E_1}^{E_2} \mathrm{d}E\, A_1(\cdots, E)A_2(\cdots, E+eV) = \int_{E_1}^{E_2} \mathrm{d}E\, A_1(\cdots, E)\frac{\partial A_2(\cdots, E+eV)}{\partial V} \tag{3.13}$$

and

$$\begin{aligned} P_2 &= A_1(\cdots, E_2)A_2(\cdots, E_2+eV)\frac{\partial E_2}{\partial V} - A_1(\cdots, E_1)A_2(\cdots, E_1+eV)\frac{\partial E_1}{\partial V} \\ &= e\left[A_1(\cdots, E_2)A_2(\cdots, E_2+eV) - A_1(\cdots, E_1)A_2(\cdots, E_1+eV)\right]. \end{aligned} \tag{3.14}$$

The conductance is therefore given by

$$G \propto \int \left(P_1 + P_2\right)\mathrm{d}\boldsymbol{k} = G_{\mathrm{Part1}} + G_{\mathrm{Part2}}. \tag{3.15}$$

A major work of this project is the comparison of the tunnelling conductance predicted by the linear and non-linear Luttinger-liquid models with experimental observations. The theoretical predictions are calculated by evaluating Equations 3.12, 3.13 and 3.14 numerically in Mathematica. The functions G_{Part1} and G_{Part2} represent the integral of P_1 and P_2 in $\boldsymbol{k}$ space, their implementation can be found in the Mathematica script for the numerical calculation. Each theoretical model is distinguished by its distinctive spectral functions. We now compare the spectral function of the different models.

2D Fermi liquid Electrons in the lower 2DEG are unaffected by the surface gates because of screening from the upper layer. They behave as a well-known example of a Fermi liquid. According to the discussion on Equation 1.45, the 2D electrons have the spectral function

$$A_2\left(k_x, k_y, \epsilon, \Gamma_2\right) = \frac{\Gamma_2}{\pi\left[\Gamma_2^2 + \left(\epsilon - \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right)\right)^2\right]}. \tag{3.16}$$

Consequently the partial derivative in Equation 3.13 is

$$\frac{\partial A_2}{\partial \epsilon} = -\frac{2\Gamma_2\left(\epsilon - \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right)\right)}{\pi\left[\Gamma_2^2 + \left(\epsilon - \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right)\right)^2\right]^2}. \tag{3.17}$$

Non-interacting 1D fermions According to Equation 1.33 and 1.42, one can see that in the absence of interactions, free fermions have sharply peaked (delta-function) spectral func-

tions centred at the free-particle energies. In reality, the behaviour of the 1D fermions in experimental samples is not ideal, as they are susceptible to scatterings by irregularities from either MBE growth or lithographic fabrication. This scattering can be considered as having similar effects to the interactions in the discussion of the Fermi liquid, which may be modelled as a convolution with a Lorentzian function of broadening parameter Γ_1. Additionally, Equation 3.8 showed that the finite confinement potential of the 1D channels leads to a transmission matrix term $\left|\tilde{\phi}(k_y)\right|^2$ in the expression of the tunnelling conductance. This term is absorbed into the 1D spectral function to give

$$A_1\left(k_x, k_y, \varepsilon, \Gamma_1\right) = \int \sum_n \delta\left(\varepsilon - \varepsilon_n\left(k_x\right) - z\right) H_n^2\left(l k_y\right) e^{-l^2 k_y^2} \frac{\Gamma_1}{\pi\left(\Gamma_1^2 + z^2\right)} \mathrm{d}z\,, \tag{3.18}$$

where $\varepsilon_n(k_x) = \hbar\omega(1/2 + n) + \hbar^2 k_x^2/2m^*$ is the energy of electrons in the n-th 1D subband, and l has the same definition as in Equation 3.7 (such that lk_y is dimensionless).

Interacting 1D fermions Equation 1.102 and 1.105 showed that according to the Luttinger liquid theory, the spectral function of 1D fermions has the form of a power-law singularity at the edge of the single-particle dispersion relation (also called the spectral edge). The spectral function of the interacting 1D fermions is obtained by replacing the delta function of the non-interacting A_1 with a power-law singularity:

$$A_1(k_x, k_y, \epsilon, \Gamma_1) = \int \sum_n \frac{\Theta\left(\epsilon - \epsilon_n\left(k_x\right) - z\right)}{\left(\epsilon - \epsilon_n\left(k_x\right) - z\right)^\alpha} H_n^2\left(l k_y\right) e^{-l^2 k_y^2} \frac{\Gamma_1}{\pi\left(\Gamma_1^2 + z^2\right)} \mathrm{d}z\,. \tag{3.19}$$

The convolution can be shown (using the analytic integration function of Mathematica) to evaluate to

$$\int_{-\infty}^{\epsilon} \frac{\Gamma}{\pi(\Gamma^2 + z^2)(\epsilon - z)^\alpha} \mathrm{d}z = \left(\varepsilon^2 + \Gamma^2\right)^{-\alpha/2} \frac{\sin\left(\frac{\alpha\pi}{2} + \alpha \tan^{-1}\left(\frac{\varepsilon}{\Gamma}\right)\right)}{\sin(\pi\alpha)}. \tag{3.20}$$

In the linear Luttinger-liquid model, the power-law exponent α in the above expressions is a constant. It can be used as a fitting parameter when we compare the calculation to the experimental results; In the non-linear Luttinger-liquid model, the power-law exponent has an expression that is dependent on the momentum of the excitation. It is given by

$$\alpha = 1 - \frac{K_c}{4}\left(1 - \frac{(k - k_\mathrm{F})\left(\frac{k_\mathrm{F}}{K_c^2} + \frac{k}{K_s^2}\right)}{t}\right)^2 - \frac{K_c}{4}\left(\frac{1}{K_c} - \frac{(k - k_\mathrm{F})\left(\frac{k_\mathrm{F}}{K_c K_s} + \frac{k}{K_c K_s}\right)}{t}\right)^2. \tag{3.21}$$

where $t = (k/K_s)^2 - (k_F/K_c)^2$, k_F is the fermi wavevector, $K_s = 1$ and K_c is used as the fitting parameter in comparison to the experiments.

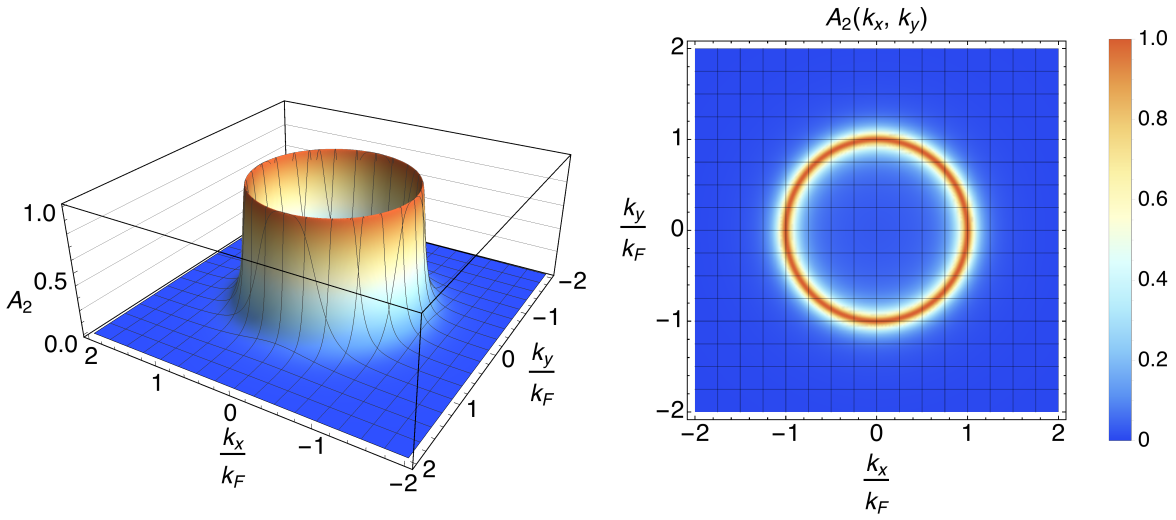
Figure 3.3 shows plots of the spectral functions discussed above, evaluated using parameters $\Gamma_1 = \Gamma_2 = 0.3\,\text{meV}$ and $\hbar\omega = 1.3\,\text{meV}$ (values typically observed in experiments) at zero V_{DC} and B, with three 1D subbands and normalised to their respective maxima. Equation 3.8 showed that the tunnelling current (and hence conductance) is proportional to the overlap of the spectral functions of the two layers. Normally, no significant overlap exists between the spectral functions of the two layers. However, by adjusting the relative offset in energy and momentum using V_{DC} and B, one can align the spectral functions to create large overlaps. Figure 3.4 demonstrates such scenarios: in the top two figures, the Fermi levels of the two layers are aligned at zero DC bias. A_1, which has a smaller radius than A_2 (since the upper layer is depleted by the wire gates), overlaps with A_2 when it is displaced by magnetic fields B_- and B_+; the lower figure shows that when a DC bias is applied, A_2 is able to 'sink' below A_1, thereby making contact with the features of A_1 which arises from the second 1D subband. Essentially, A_2 acts as a probe of A_1, which is moved by sweeping B and V_{DC}. One can therefore build up a picture of the convolution of the spectral functions over arbitrary ranges of energy and momentum by measuring the tunnelling conductance.

3.2 Maximum of tunnelling conductance

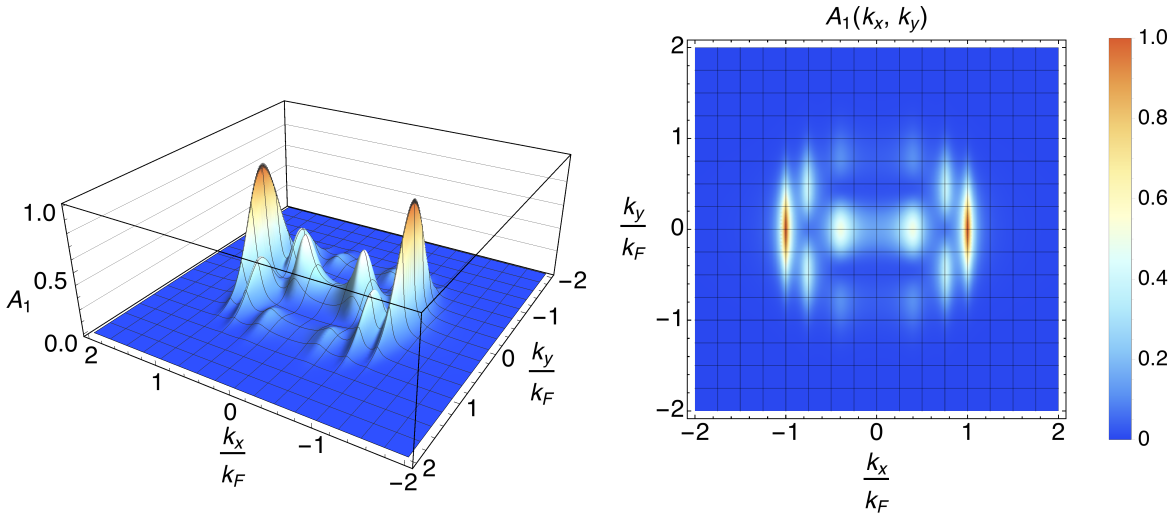
An important task of analysing the tunnelling data is calculating the position of the maximum conductance in the V_{DC}–B space and fitting it to the measured result. The previous section showed that a necessary condition for tunnelling is that the initial and final states must have the same momentum and energy. Furthermore, Equation 3.16, 3.18 and 3.19 showed that all the spectral functions considered in this work are peaked around the free-particle energies. Since the tunnelling conductance is proportional to the convolution between the spectral functions of the two layers, we expect the maximum to occur when the empty states of one layer line up with filled states of another layer in the free-particle dispersion. The positions in V_{DC}–B space where such alignments occur are given by the equations

$$\begin{aligned} \text{Low/High1D} \qquad & eV_{DC} = \frac{\hbar^2}{2m^*}\left[k_{F1}^2 - \left(k_{F2} \pm \frac{eBd}{\hbar}\right)^2\right] \\ \text{Low/High2D} \qquad & eV_{DC} = \frac{\hbar^2}{2m^*}\left[\left(k_{F1} \pm \frac{eBd}{\hbar}\right)^2 - k_{F2}^2\right], \end{aligned} \tag{3.22}$$

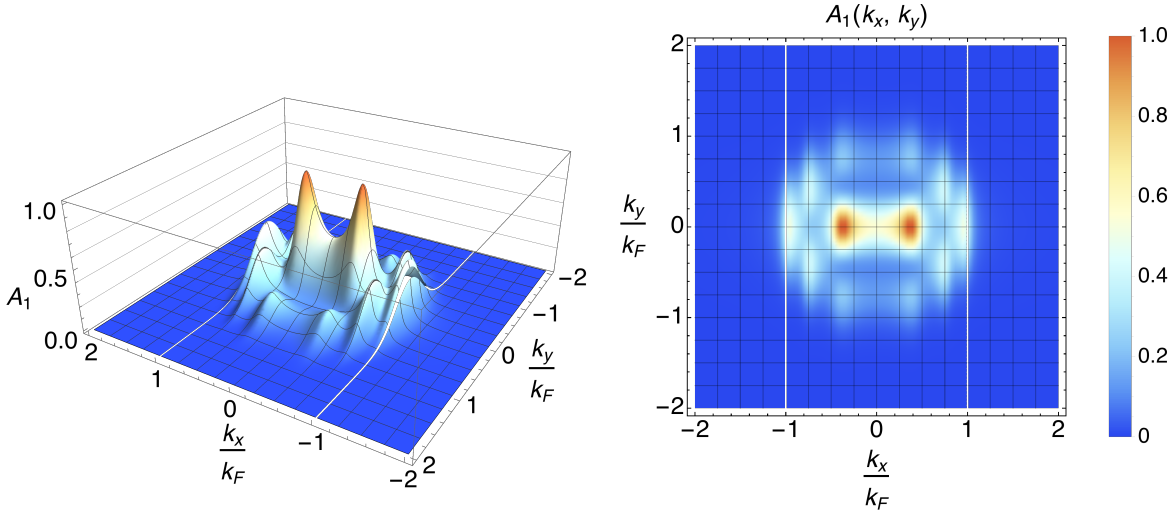
where k_{F1} and k_{F2} are the Fermi wavevectors of electrons in the 1D/2D layer, and m^* the electron effective mass in GaAs. The names of the equations follow the function definitions



(a) The 2D Fermi-liquid spectral function.



(b) The 1D non-interacting spectral function.



(c) The 1D linear Luttinger-liquid spectral function.

FIGURE 3.3 Spectral functions evaluated at $eV = 0$ and $B = 0$ with parameters: $\Gamma = 0.3$ meV and $\hbar\omega = 1.3$ meV using Mathematica. k_x and k_y are expressed in units of the Fermi wavevector k_F. Each figure is normalised to their respective maximum value.

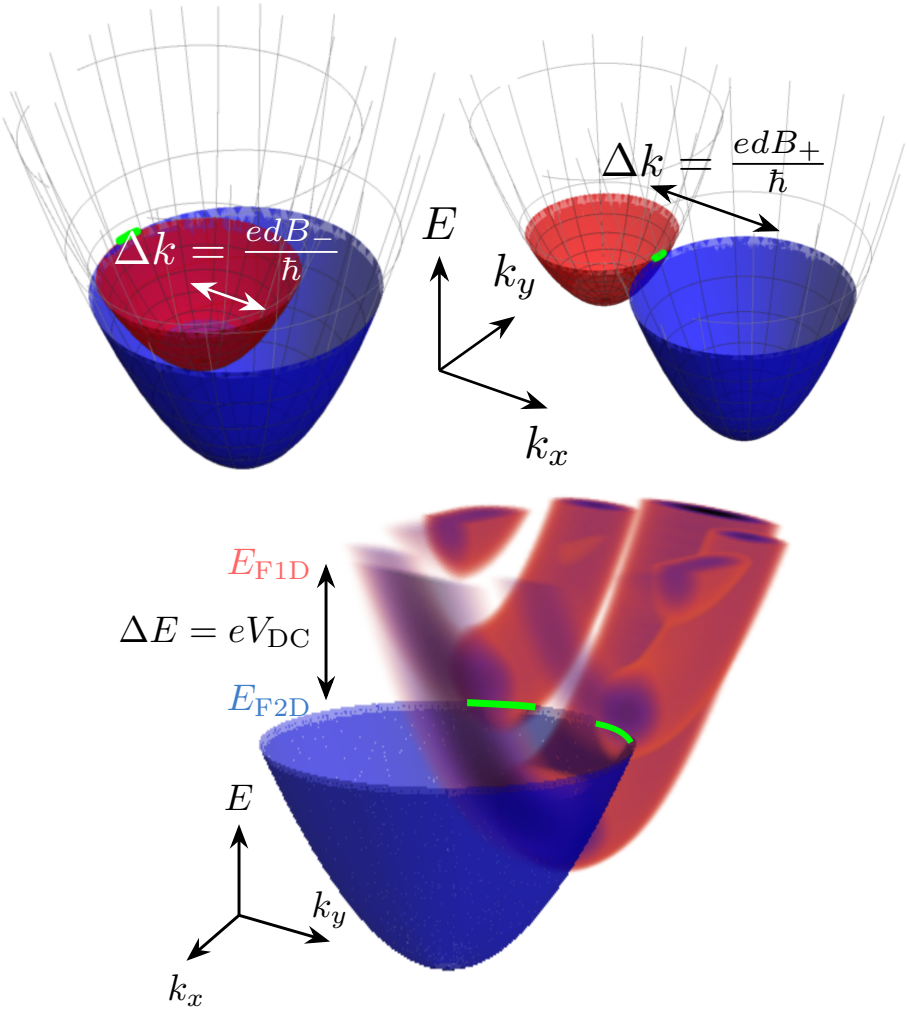
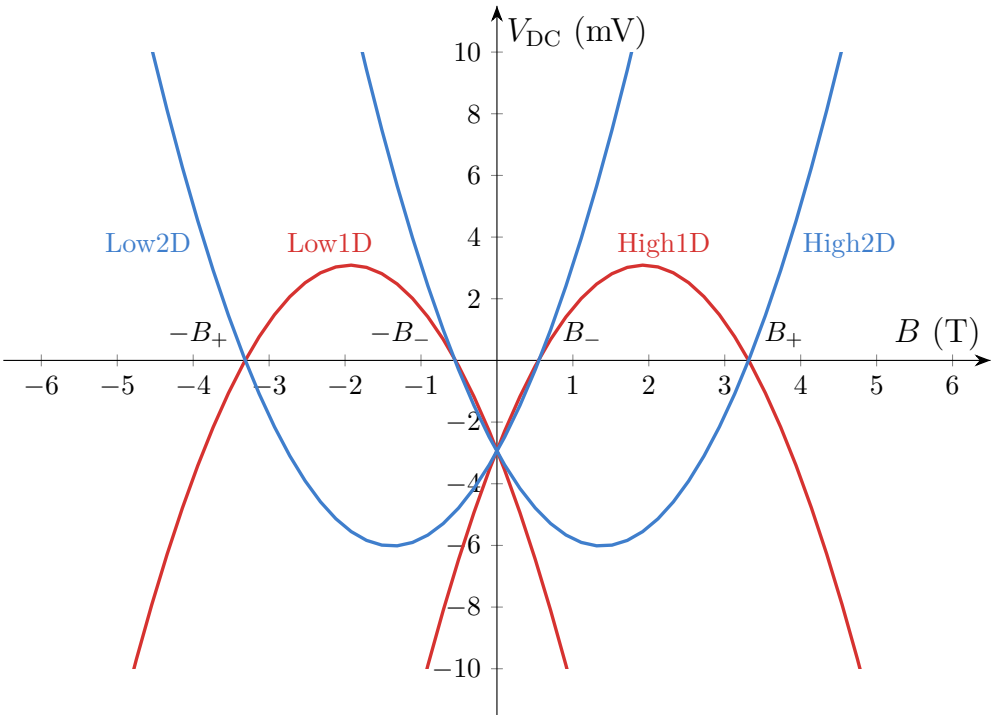
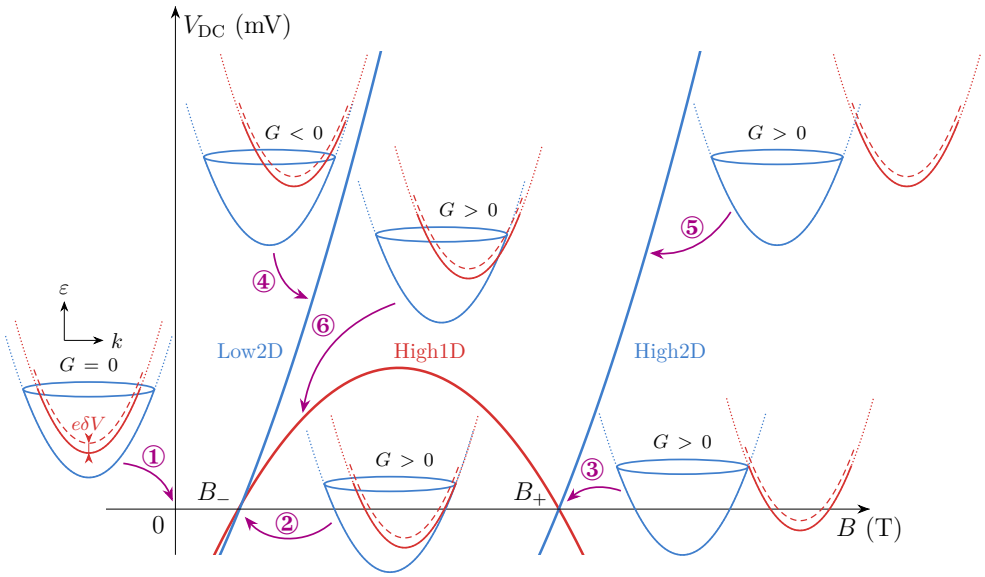


FIGURE 3.4 The 1D (red) and 2D (blue) spectral functions in energy-momentum space. Momentum offset is induced by the magnetic field: $\hbar\Delta k = edB/\hbar$; energy offset is induced by the DC bias: $\Delta E = eV_{\mathrm{DC}}$. At any instant, the tunnelling current is proportional to the overlap of the spectral functions, marked by the green line segments. Adapted from plots by Prof. C. J. B. Ford.

in the MATLAB scripts for result analysis. Physically, the '1D' equations correspond to the dispersions of electrons in the upper layer, and their '2D' counterparts correspond to the dispersions of electrons in the lower layer. The shapes of the parabolae are determined by the parameters k_{F1}, k_{F2}. As will be discussed shortly, the magnetic fields at which the parabolae cross the the zero V_{DC} line, B_- and B_+, can be used to calculate k_{F1} and k_{F2}. Figure 3.5a shows a set of curves plotted in a typical experimental range of V_{DC}–B, with k_{F1}, k_{F2} calculated from crossing points observed from wafer C-2617, and $m^* = 0.067m_{\mathrm{e}}$, $d = 32\,\mathrm{nm}$. Figure 3.5b shows the same plot zoomed-in to the first quadrant of the V_{DC}–B space, with six insets pointing to different parts of the curves to demonstrate the alignment of the 1D/2D



(a) Maximum conductance curves given by Equation 3.22. The set of curves shown here have parameters $k_{F1} = 7.36 \times 10^7$ m^{-1} and $k_{F2} = 1.03 \times 10^8$ m^{-1}, calculated from experimentally observed crossing points of wafer C-2617 at $B_- = 0.55$ T and $B_+ = 3.32$ T; $m^* = 0.067 m_e$, the standard effective electron mass in GaAs.



(b) Tunnelling conductance in the first quadrant of V_{DC}–B space and the associated 1D/2D dispersion alignments in the energy–momentum space. The 1D dispersions are represented by the red parabolae and the 2D ones the blue paraboloids. The solid part of the dispersions represents the filled states and the dotted part the empty states. The dashed parabolic profile above the 1D dispersions represents the deviation of the filled states caused by a small, positive AC bias δV, which is used for the phase-sensitive measurement technique.

FIGURE 3.5 Curves of maximum tunnelling conductance in the V_{DC}–B space.

dispersions there. A dashed profile is drawn on top of each 1D dispersion to show the deviation from their equilibrium position caused by a small, positive AC bias δV. Application of the AC bias is part of the phase-sensitive measurement technique, to be discussed in Section 6.1.2. Most importantly, it detects the differential tunnelling conductance $G = \delta I / \delta V$, where δI is the change in the tunnelling current in response to δV. Inset ① shows that at the origin, where the energy and momentum offsets between the layers are zero, the filled states on both dispersions are matched at their Fermi levels and are concentric in k space. Since there is no overlap between any filled and empty states, the tunnelling current and conductance are zero. Inset ② and ③ respectively show that at the crossing points $\boldsymbol{B}_-$ and $\boldsymbol{B}_+$, the 1D Fermi points line up with the 2D Fermi surface internally and externally. At the equilibrium position (defined by the DC level), the filled states in both layers are aligned and therefore the tunnelling current is zero. However when the 1D dispersion is raised by δV, the filled 1D states become aligned with the empty states just above the 2D Fermi surface, and tunnelling occurs. The increased tunnel current in response to δV suggests the differential conductance is positive at the crossing points. Inset ④ and ⑤ follow Low2D and High2D respectively in the positive V_{DC} regime. They show that these two curves are formed by the 1D Fermi points tracing out the empty states of the 2D dispersion. Furthermore, ④ shows that the differential conductance along Low2D is negative at large positive V_{DC}: at the equilibrium position, the filled states close to the right 1D Fermi point align with empty 2D states, which results in positive tunnelling current (in the sense that electrons flow from upper to lower layer). When the 1D dispersion is raised by δV, it moves out of alignment with the 2D dispersion completely, which causes the tunnelling current to drop. The differential conductance is therefore negative. On the contrary, ⑤ shows the conductance on High2D is positive. This is because the current is initially zero when the 1D dispersion lies just below the 2D dispersion, but when raised by the AC it becomes positive. Finally, ⑥ shows that High1D is traced out by the empty states just above the 2D Fermi surface, following the filled states on the 1D dispersion. The conductance is positive for the same reason as ⑤.

In order to calculate k_{F1} and k_{F2} from the zero DC crossing points, consider Figure 3.6, which illustrates the relative offsets between the spectral functions in k-space at $\boldsymbol{B}_-$ and $\boldsymbol{B}_+$. It shows

$$\Delta k_- = k_{\mathrm{F2}} - k_{\mathrm{F1}} = \frac{ed}{\hbar}\boldsymbol{B}_- \qquad \Delta k_+ = k_{\mathrm{F2}} + k_{\mathrm{F1}} = \frac{ed}{\hbar}\boldsymbol{B}_+, \tag{3.23}$$

where Δk is the wavevector offset between the two layers. These can be solved for the Fermi wavevectors

$$k_{\mathrm{F1}} = \frac{ed}{2\hbar}(\boldsymbol{B}_+ - \boldsymbol{B}_-) \qquad k_{\mathrm{F2}} = \frac{ed}{2\hbar}(\boldsymbol{B}_+ + \boldsymbol{B}_-). \tag{3.24}$$

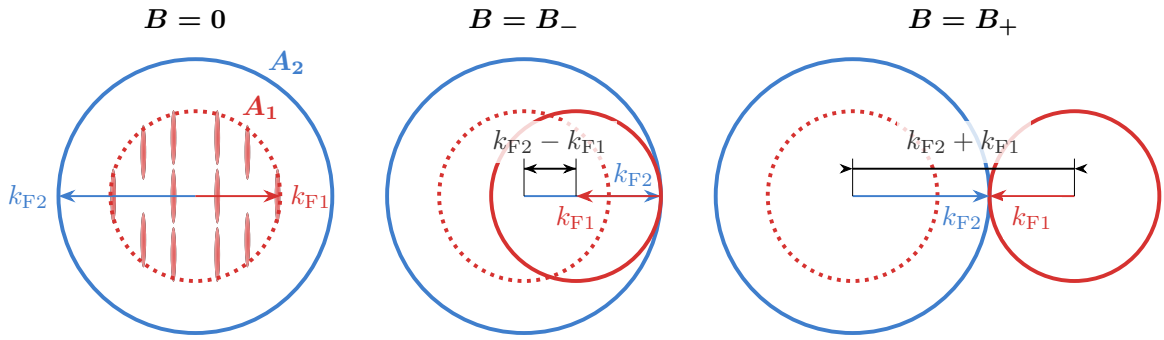


FIGURE 3.6 The relative positions of spectral functions in the momentum space at zero magnetic field and the two zero-DC crossing points. The blue circle represents the 2D spectral function; the red circle represents the approximate extent of the 1D spectral function, the oval shapes in the left plot represent the actual spectral function, see Figure 3.3 for the detailed plot. The black arrow labels displacement in k required to bring the two spectral functions into contact from zero field.

Having obtained the k_F values, we can calculate the electron densities in the 1D/2D layers with the standard density-of-states expressions of free-electron states

$$n_{1D} = \frac{dN}{dL} = \frac{2k_{F1}}{\pi} \qquad n_{2D} = \frac{dN}{dA} = \frac{k_{F2}^2}{2\pi}. \tag{3.25}$$

When the maximum conductance curve is plotted on top of experimental data, one finds that the curves do not line up well with the actual maxima. The calculation based on the free-particle dispersion must be modified to take account of additional effects such as the capacitive charging.

3.2.1 Capacitance correction to the conductance curves

The two closely-spaced layers of parallel 2DEGs form a capacitor which becomes charged when a DC bias is applied across them, and hence the carrier densities in the two layers are changed. This causes the position of the observed conductance maxima to deviate from the free-particle dispersions, which is more pronounced the higher the DC voltage. Suppose at DC bias V_{DC}, the change in the total electric charge in the 1D/2D layer due to the capacitive effect is

$$\delta Q_i = \eta_i \Gamma V_{DC}, \tag{3.26}$$

where the subscript $i = 1, 2$ labels the 1D and 2D layer, respectively, Γ is the capacitance, and $\eta_i = \pm 1$ is a sign factor based on the experimental setup. Since the positive pole of the DC bias is connected to the 2D layer in our experiment, a positive V_{DC} will cause n_{2D} to fall and n_{1D} to increase, therefore $\eta_1 = 1, \eta_2 = -1$. Changes in the carrier densities n_i associated

with the change in the total charge is

$$\delta n_i = \frac{\delta Q_i}{eD_i}, \tag{3.27}$$

where $D_1 = L, D_2 = A$ are the physical dimensions (length and area) of the 1D/2D system. Combining Equations 3.25–3.27, we have:

$$\begin{aligned} \delta n_1 = n_1' - n_1 = \frac{\Gamma V_{\mathrm{DC}}}{eL} = \frac{2}{\pi}\left(k_{\mathrm{F1}}' - k_{\mathrm{F1}}\right) \\ \delta n_2 = n_2' - n_2 = -\frac{\Gamma V_{\mathrm{DC}}}{eA} = \frac{1}{2\pi}\left(k_{\mathrm{F2}}'^2 - k_{\mathrm{F2}}^2\right), \end{aligned} \tag{3.28}$$

where the primed quantities are modified by the capacitive effect and the unprimed ones the zero-DC counterparts. The modified Fermi wavevectors at non-zero DC bias are

$$k_{\mathrm{F1}}'(V_{\mathrm{DC}}) = k_{\mathrm{F1}} + \frac{\pi\Gamma V_{\mathrm{DC}}}{2eL} \qquad k_{\mathrm{F2}}'(V_{\mathrm{DC}}) = \sqrt{k_{\mathrm{F2}}^2 - \frac{2\pi\Gamma V_{\mathrm{DC}}}{eA}}. \tag{3.29}$$

By substituting k_{F1} and k_{F2} in Equation 3.22 with their primed counterpart given by Equation 3.29, one obtains the expressions of the capacitance-corrected positions of the conductance maximum. e. g. Low1D becomes

$$eV_{\mathrm{DC}} = \frac{\hbar^2}{2m^*}\left[\left(k_{\mathrm{F1}} + \frac{\pi\Gamma V_{\mathrm{DC}}}{2eL}\right)^2 - \left(\sqrt{k_{\mathrm{F2}}^2 - \frac{2\pi\Gamma V_{\mathrm{DC}}}{eA}} + \frac{eBd}{\hbar}\right)^2\right] \tag{3.30}$$

This needs to be solved for V_{DC} in terms of B and Γ. The solution obtained with Mathematica (using the `Solve` function) is too long to be reproduced here. They can be found in the definitions of the functions `Low1p`, `High1p`, `Low2p` and `High2p` in the MATLAB subroutine `plot_parabolae.m`, used to plot the dispersions with the data. Two additional parameters appear in the modified dispersion formulae: Γ/L and Γ/A–the capacitance per unit length and area, of the upper and lower layer. Their values are determined iteratively during analysis of the result, by varying Γ/A until the calculated curves fit the data as well as possible. The value of Γ/L is deduced from Γ/A: we assume that the two layers have the same capacitance per unit area, and convert the per-unit-area capacitance of the upper layer into the per-unit-length equivalent by noting that from Equation 3.30, the ratio between Γ/L and Γ/A is

$$\frac{\Gamma/L}{\Gamma/A} = \frac{\frac{2}{\pi}(k_{\mathrm{F}}' - k_{\mathrm{F}})}{\frac{1}{2\pi}\left(k_{\mathrm{F}}'^2 - k_{\mathrm{F}}^2\right)} = \frac{\frac{2}{\pi}\delta k_{\mathrm{F}}}{\frac{1}{2\pi}\left(k_{\mathrm{F}}\delta k_{\mathrm{F}} + \delta k_{\mathrm{F}}^2\right)} = \frac{4}{k_{\mathrm{F}}}. \tag{3.31}$$

The subscripts of the k_F have been suppressed, because in this case they are of the same layer. δk_F represents the small change in Fermi energy caused by the DC voltage.

3.3 Summary

In this chapter, we have derived the formulae for calculating the current and conductance in our resonant tunnelling device. The conclusion of this derivation is that both quantities are proportional to the convolution between the spectral functions of the two conducting layers of the substrate. By applying a DC bias and a magnetic field, one can induce arbitrary offsets along the energy and momentum axes in the spectral functions, thereby constructing a mapping of the convoluted spectral functions in the full energy-momentum space. This finding is significant and forms the experimental foundation of our work, because it enables us to calculate the tunnelling conductance if we can find out the spectral functions of the two layers. Indeed, the core study of this work is based on the comparison between the computer modelled tunnelling conductance based on different theories and the experimental data gathered from physical measurement. In the previous chapter, we have seen that different theoretical models of the 1D electrons give rise to distinctive spectral functions. By substituting them into the conductance formula found in this chapter, we obtain calculate different predictions for the tunnelling conductance. We can compare these against the actual experimental data of tunnelling conductance and discover which theory better models the experiment. The details of this study will be covered in the data analysis chapter later.

In addition to the discussion on the calculation of the tunnelling conductance across a whole range of energy and momentum, we have also studied how to determine the individual subbands by using the maxima of the conductance. We have seen that to improve the accuracy of the maxima calculation, one must modify the free-particle dispersion with corrections due to capacitive effects. This will prove to be another crucial step in the analysis of data, as we will soon see in later chapters that not only will the subband position provide us with a range of important parameters about electrons in the two layers, but more importantly, its position reveals the spectral edge, which serves as a reference to our data fitting works.

Chapter 4

Sample fabrication and processing

The aim of this project is to study the effects which arise from electron-electron interactions in one-dimensional geometry. To this end, intricate semiconductor devices were designed and fabricated, which enabled us to confine electrons in narrow channels and measure their momentum and energy response to various excitations. The devices were built on GaAs/AlGaAs wafers, a material commonly used for high-electron-mobility transistors. Two parallel layers of free electrons were built into our wafers by MBE growth, separated by a tunnel barrier. The electrons can tunnel between the layers when specific conditions are met in terms of the their energy and momentum. This tunnelling event can therefore be used as a spectrometer which reveals the energy and momentum of the conducting electrons in each layer. This chapter describes the techniques employed to process plain semiconductor wafers into functional experimental devices.

4.1 Wafer material

All experiments reported in this work were carried out on devices fabricated on modulation-doped GaAs/$Al_xGa_{1-x}As$ ($x = 0.33$) heterostructures, grown with molecular-beam epitaxy (MBE) in the Cavendish Laboratory. Most experimental results reported in this work were performed on the wafers C2617, W938 and W939. The C2617 wafer was originally used by Dr. Yodchay Jompol for his doctoral work [44]. W938 and W939 were replicas of C2617 made by Dr. Ian Farrer with minor modifications. The composition of wafer C2617, listed from the surface to the substrate, is shown in Table 4.1. A model of the conduction band structure is shown in Figure 4.1. W938 and C2617 have identical structures, while W939 differs from W938 by having a 100×2.5 nm/2.5 nm GaAs/AlGaAs super-lattice below the 350 nm AlGaAs under the lower quantum well. Both wafers feature two parallel quantum wells (QW) separated by a tunnel barrier of GaAs/AlGaAs superlattice, which increases the spacial separation

TABLE 4.1 MBE layers of wafer C2617 from surface to substrate

Material	Thickness (nm)	Function
GaAs	16	Protective cap
AlGaAs	40	Doping layer with 1×10^{18} cm^{-3} of Si
AlGaAs	20	Modulation doping spacer
GaAs	18	Upper quantum well
GaAs/AlGaAs	$10 \times (0.556/0.833)$	Superlattice tunnel barrier
GaAs	18	Lower quantum well
AlGaAs	25	Modulation doping spacer
GaAs	0.56	Spacer
AlGaAs	15	Spacer
AlGaAs	40	Doping layer with 1×10^{18} cm^{-3} of Si
GaAs	0.56	Spacer
AlGaAs	350	Transition to bulk

of the two quantum wells, hence preventing the hybridisation of the electron wave functions in the two QWs, while maintaining a low enough average barrier height to ease tunnelling.

4.2 Optical processing

Sample fabrication begins with the physical shaping of MBE wafers: using a scribing station, a wafer is cut into small, rectangular chips. The chips are processed in a series of optical lithography stages, which eventually shape the surface of the chips into Hall bars and electrical contacts of metal be deposited onto the surface. Collectively, these processing steps are referred to as the optical stages.

Optical lithography is a process in which patterns on a pre-manufactured template (called the mask) are transferred onto the chips using ultraviolet (UV) radiation. To start, a sample is coated in thin layers of chemicals called photoresists, and then exposed to UV radiation under the mask. The mask is a piece of quartz glass which is coated in a layer of chromium that features patterns of openings which allow the UV radiation to transmit through the glass. Depending on the type of photoresist, the UV radiation results in chemical changes which make it either soluble or insoluble to a special solvent called the developer. The photoresist is said to be positive if the part exposed to UV becomes soluble, and negative if it is insoluble. By immersing the sample in a bath of developer, the soluble portion of the photoresist is stripped away, leaving a chip covered in a layer of patterned photoresist. Optical lithography

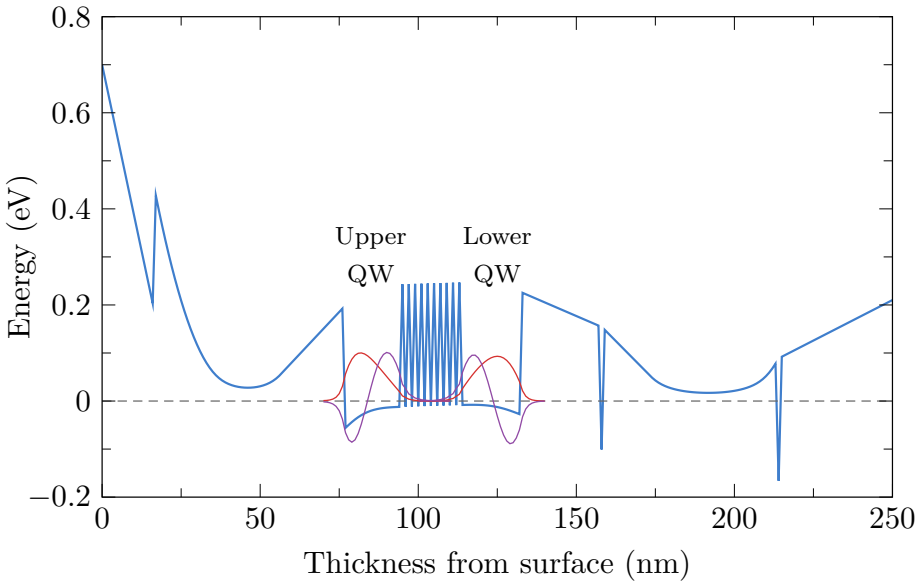


FIGURE 4.1 Conduction band energy profile around the quantum wells of wafer C2617 and the first two eigenstate states of electrons in the wells, as calculated by the 1D Poisson solver by G. Snider. The calculation was run with the parameters: (a) 0.7 eV Schottky barrier at the surface. (b) Fully-ionised donors.

is repeated three times for the processes of chemical etching of Hall bars, fabrication of Ohmic contacts and deposition of Schottky contacts. All of the processes were carried out with positive-tone resists with a mask custom-made by specialist company JD Photo-Tools.

4.2.1 Etching of Hall bars

The optimal chip dimension is 11 mm × 12 mm, which makes full use of the 5 × 5 device pattern on the photo-mask, as well as allowing for clearance at the corners of the chip for edge-bead removal. Prior to photolithography, the chips were cleaned in acetone then IPA (isopropyl alcohol) in a bath with ultrasonic agitation, then dried under a nitrogen blowgun. They were then baked on a 155 °C hotplate for five minutes to dry out the solvents. The chips were coated with MicroChem S1813 photoresist using a resist spinner rotating at 5500 rpm for 60 s to achieve a thickness of approximately 150 nm, and baked for 60 s on a 125 °C hotplate to harden the resist. Afterwards, a Hall bar pattern was exposed onto the sample with a mask aligner. Care must be taken during alignment to ensure that the Hall bars are parallel to the high-mobility direction of the wafer (also referred to as the major flat cleavage plane for GaAs). After the UV exposure, the chips were developed at room temperature in MicroChem MF-319 developer for 30 s, then rinsed in de-ionised (DI) water and nitrogen blow-dried.

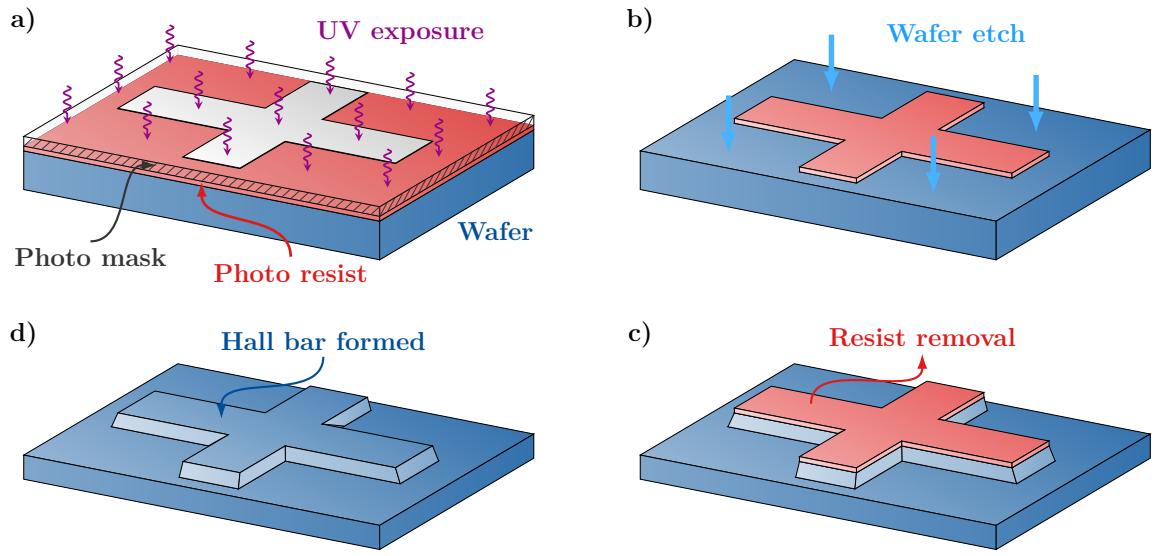


FIGURE 4.2 The fabrication of a Hall bar by photolithography. a) Exposure of photoresist. b) Etching of wafer after resist development. c)-d) The resist is removed after etching, revealing the Hall bar.

Hall bars are formed in a wet-etching process, where the developed chips are immersed in a mixture of water, sulphuric acid and hydrogen peroxide. The chip surface covered by photoresist is protected from the chemicals while the exposed material reacts and dissolves in the etchant, leaving behind an elevated structure called the mesa, which forms the Hall bar. The volume ratio of the etchant used in our processing was 1000:8:1 of H_2O:H_2SO_4:H_2O_2, which had the property of leaving an over-cut cross-sectional profile normal to the high mobility axis. The over-cut profile is the preferable finish, particularly compared to an under-cut profile, because its positive incline is conducive to forming a continuous layer of evaporated metal contacts (which is the goal of the next processing step). An under-cut profile is liable to cause metal contacts to break when they try to surmount its negative incline.

According to table 4.1, the lower QW is 125 nm beneath the wafer surface. A chip must be etched deeper than this figure to have the Hall bar contain both conducting layers. Before the chips were immersed in the etchant, we measured the thickness of the developed resist with a DekTak surface profiler. The chips were then etched by 300 nm in iterations of short steps (for every 90 seconds each time) and DekTak measurements. The total etch time was around 6 minutes at cleanroom temperature (18 to 20 °C). After the target depth was achieved, the sample was immersed in acetone to dissolve away the photoresist. It was then rinsed in IPA and dried with nitrogen blowgun ready for the next process.

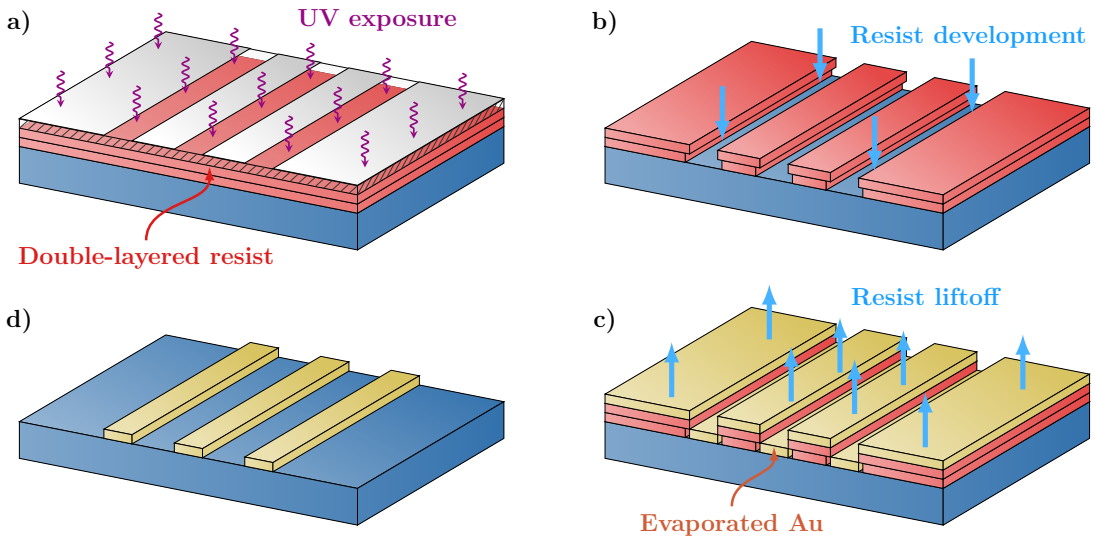


FIGURE 4.3 The fabrication of surface gates by positive-tone photolithography. a) Exposure of photoresist. b) Development of double resist. Note the choice of resists should result in an over-cut cross-sectional profile. c-d) Metallisation by thermal evaporation and lift-off.

4.2.2 Ohmic contacts

In order to pass electric current through the conducting layers of the chips, contacts must be made between the surface of the wafer and the QWs which are 125 nm underneath. This is achieved by diffusing an alloy of gold-germanium-nickel (AuGeNi) into the GaAs/AlGaAs crystal under intense heat in a process called annealing. The metal-semiconductor junction thus formed is a non-rectifying junction which has linear current-voltage response, known as an ohmic contact. [86]

Ohmic contacts were fabricated at the far ends of Hall bars to be used as current sources and sinks. Photolithography was used to deposit AuGeNi at the desired positions on the Hall-bars. Two layers of photoresists are spun onto to the sample in sequence: the first layer differs from the second by being more sensitive to UV radiation. The exposure and development proceed in the same manner as the Hall bar process. During development, the higher sensitivity of the lower resist layer results in higher rate of dissolution than the upper layer. Consequently, the developed double-layer resists have an under-cut cross-sectional finish. The AuGeNi alloy is deposited as a 150 nm-200 nm film over the developed resist in a high-vacuum thermal evaporator. Afterwards, the sample is immersed in SVC14 resist remover. As the resists dissolve, the metal film which lay on top of the resists is separated from the sample. Only the portion that was in direct contact with the sample surface remains. This is referred to as the lift-off process. As explained in the previous section, the under-cut profile of the double-layer resist results in discontinuity in the metal film which makes it easier to

separate it from the sample. After lift-off, the sample was heated in a rapid annealing oven, which was used to quickly heat the sample (so as to minimise heat damage to the sample) up to 500 °C for about 10 s, to diffuse the alloy film into the substrate.

In practice, MicroChem LOR5B and S1805 photoresists were used to form the lower and upper layers, respectively. LOR5B was applied in a resist spinner with a speed of 6000 rpm for 60 s, followed by a 5-minute bake on a 180 °C hotplate. S1805 was applied with a speed of 5500 rpm for 60 s, followed by a 5-minute 115 °C hotplate bake. Before AuGeNi evaporation, the sample was dipped in 20% hydrochloric acid (HCl) for 10 s in order to remove the oxide on the substrate surface, a crucial step which improves the electrical properties of the finished ohmic contact. Lift off was carried out in a petri dish of SVC14 resist remover heated to 90 °C on a hotplate and lasted for 15 minutes, followed by a 1-minute ultrasonic bath in IPA. Annealing was carried out in an automated oven, which ran a program with a peak temperature of 430 °C for 80 s. Once annealed, the ohmic contacts were tested at room temperature on a probe station, which measures the electrical resistance between every pair of contacts. On wafer W938 and W939, each pair of ohmic contacts registered $\sim 5500\,\Omega$ of resistance. A subtle design feature was that ohmic metal was also deposited at positions off the Hall bar, where the optical gates (fabricated in the next process) would be. This was done in order to increase the thickness of bond pads (thereby improving wire-pad adhesion in the wire-bonding process in the final stage of fabrication). If the Hall bar was not etched deep enough past the lower QW, these extra ohmic contacts would be connected to the ohmic contacts on the Hall bar via the two QWs. To verify that such leakage path does not exist, the resistance between the bonding pads and the ohmic contacts should be $\sim 4.5\,\mathrm{M}\Omega$ at room temperature without light on the probes.

4.2.3 Surface gates

Surface gates differ from the ohmic contacts by being current-rectifying metal-semiconductor junctions, also known as Schottky barriers. On our devices, they are formed by gold (Au) contacts, which consist of bonding pads and tracks on the surface of the sample. They serve as the intermediary connection between the sample package and the substantially smaller gates fabricated by e-beam lithography. They were made using the same photo-lithographic technique as the ohmic contacts, i. e. the bi-layer photolithography process involving LOR5B and S1905 resists. It differs from the ohmic contact process in the step of metallisation: for the surface gates, a layer of titanium was first evaporated onto the substrate, followed by a layer of gold. The Ti layer acts as a wetting layer, which adheres strongly to both the Au layer and the GaAs/AlGaAs surface. (Au does not adhere strongly to GaAs/AlGaAs.) It also acts as a 'getter' during thermal evaporation by actively binding to impurities in the evaporation

chamber, which lowers the ambient pressure in the evaporator, thus improving the quality of the finished contact. Due to its high melting point, care must be taken to limit the exposure time of the sample to the Ti crucible, as heat from prolonged exposure can cause the photoresist to melt and re-flow. On the other hand, Au has a much lower melting point. It is preferable to apply a thick layer of Au so that the bonding pads can withstand the subsequent wire bonding process without breaking. For the device reported in this work, 5 nm of Ti and between 120 to 150 nm of Au were applied. As for the ohmic contact process, the quality of surface gates can be improved by removing the surface oxide and hence any residual resist on top of it with HCl prior to Ti/Au evaporation. Without this step, the surface contacts can suffer from poor adhesion to the substrate surface. The lift-off process of the optical gates is identical to that of the ohmic-contact process.

4.3 Electron-beam lithography

Optical lithography is limited in the dimension of patterns it can produce by the wavelength of the UV radiation. Additionally, the accuracy of the manual mask aligner is limited by the operator's experience, which leads to alignment error between different sets of optical patterns. Consequently the lower limit of pattern dimension is typically ∼0.8 μm. Features smaller than this limit can be fabricated using the method of electron-beam lithography (EBL), which uses electrons from a high-voltage electron gun, instead of UV radiation, to break down long-chain resist polymer molecules. The Semiconductor Physics Group in the Cavendish laboratory uses a fully automatic Vistec VB6 UHR EWF system for EBL. Typically operating at 50 or 100 kV with a 4 nm spot size, the system employs a laser interferometer stage with an alignment precision of 0.62 nm. The principle of EBL is similar to that of photolithography: an electron beam (e-beam) is focused onto a resist-coated sample. By the combined action of magnetic deflection coils and the motorised stage tracked by laser-interferometer, the e-beam raster scans the sample surface according to patterns programmed into the control computer, breaking up the long molecular chains of the resist polymer (typically poly-methyl methacrylate, or PMMA) into smaller segments, which become soluble in developer. The primary sub-micron features of the tunnelling device are arrays of intricate parallel surface gates, plus a series of three-dimensional structures called air-bridges which inter-connects the surface gates at their two ends. The fabrication of these features requires two stages of EBL: in the first stage, a 2D pattern is produced over the sample surface which forms the array; then the second stage produces the three-dimensional air-bridges, which span across the gaps of the base layer and interconnect the individual gates.

4.3.1 Base layer

The standard EBL process that was used to fabricate the gate arrays was similar to the photolithography process described in the previous sections. First, a double-layer resist was applied to the sample: the lower layer was 125 nm of 100k molecular-weight A6 (6% solid content in anisole) PMMA, deposited by a resist spinner at 8000 rpm for 60 s. The upper layer was 100 nm of 950k A4 PMMA, diluted in 2:1 ratio to methyl isobutyl ketone (MIBK), also deposited at 8000 spinner rpm for 60 s. After each layer was applied, the sample was baked on a 125 °C hotplate for three minutes. The gate pattern was exposed by a 50 kV e-beam at dose 550 μC cm^{-2}. Development of the exposed sample was carried out in a mixture of 1:3 MIBK/IPA for 30 s at room temperature. Because of the lower molecular weight, the base-layer resist has a faster rate of development than the upper layer, hence the double-layer structure leads to an undercut cross-sectional profile after development. Oxide removal was performed in 20% HCl for 10 s prior to metallisation. The surface gate was formed of a thermally evaporated film of 5 nm Ti and 20 nm Au. Lift-off was performed by immersing the sample in acetone overnight before manually agitating the metal film with a pipette. Ultrasonic bath should be avoided on EBL metallisations to avoid damaging the delicate gates.

4.4 Air-bridge process

4.4.1 Overview

The fabrication of ultra-fine three-dimensional interconnections of the surface-gate array required detailed study, as no similar EBL process had been attempted in the Semiconductor Physics Group prior to this project. A variety of methods had been reported in literature for making microscopic three-dimensional structures. One approach is to use a spacer layer of insulating material, such as regular [26] or cross-linked PMMA [100, 5, 34] to elevate the 3D structure. Alternatively, it can be made to arch over any underlying structures in free space. For this project, the latter approach is preferable for the following reason: the 3D structure and the 2DEG under the wafer surface can be considered as a parallel-plate capacitor with two series layers of dielectric—the spacer and the GaAs/AlGaAs substrate, whose capacitance can be written as

$$C = \frac{1}{1/C_{\mathrm{sp}} + 1/C_{\mathrm{sub}}} = \frac{\varepsilon_0 A}{d_{\mathrm{sp}}/\varepsilon_{\mathrm{sp}} + d_{\mathrm{sub}}/\varepsilon_{\mathrm{sub}}}, \tag{4.1}$$

where ε is the dielectric constant and d the thickness of dielectric. By using an air-gap instead of a layer of any other dielectric material (such as PMMA), $\varepsilon_{\mathrm{sp}}$ is minimised and consequently so is the capacitance C between the bridge and the 2DEG.

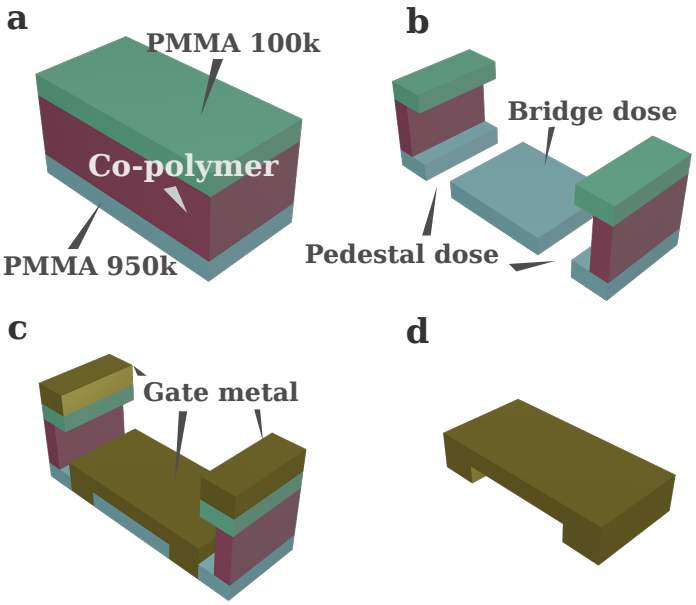


FIGURE 4.4 EBL process for air bridge. a) A triple-layer stack of resist is deposited onto the sample surface. b) The resist stack is exposed by two different EBL doses, and the figure shows the structure of the stack after development. c) Gate metal is thermally evaporated onto the developed resists. d) After liftoff, the part of the metal which is in contact with the sample (pedestals) remains.

Several methods have been reported on how an air-gap geometry can be achieved: [58] demonstrated a process involving nanoimprint lithography (NIL), where a pre-fabricated template was used to impress the desired three-dimensional feature onto a layer of resist. This method is capable of producing large numbers of ultra-fine features with very high precision, but it lacks flexibility—any template is specific to a particular pattern and is not suitable to be used in projects such as our work, where rapid prototyping is required. An alternative approach was to directly shape the resist with e-beam. [30] reported a method where a single layer of PMMA was exposed by an e-beam with variable-voltage and [53] demonstrated a method with a multi-layer resist of PMMA and polyamide. These two approaches satisfy the requirements for rapid prototyping but are unsuitable for ultra-fine features: In the first method, changing the e-beam voltage during exposure requires the EBL machine to be re-calibrated due to focus change. Lowering the beam voltage also reduces its resolution. The second method does not suffer from these drawbacks, but does require two EBL exposures due to the use of polyamide.

Having considered the advantages of several techniques for 3D fabrication in the past, we developed a variable-dose, triple-PMMA-resist EBL process which combines the flexibility of rapid prototyping with high-resolution patterning. Additionally, the process requires only a single stage of exposure, making it easy to implement. The three layers of resist that the process uses are: a bottom layer of 150 nm of 950k PMMA, a middle layer of 300 nm of methyl

methacrylate (MMA)(8.5)-methacrylic acid (MAA) co-polymer and a top layer of 130 nm of 100k PMMA. The triple-layer stack is exposed with two distinct e-beam doses: the first is a high dose (referred to as the pedestal dose) with enough energy to fully expose all three layers of resist while the second is a low dose (referred to as the bridge dose) which can only expose the top and middle layer, leaving the bottom layer unaffected and insoluble to the developer. Figure 4.4 is an illustration of the flow of the air-bridge process and Figure 4.5 shows SEM images of air-bridge arrays on a working device.

The greatest challenge of the air-bridge process was in ensuring good adhesion between the air-bridges and the underlying metal. Although one may expect the air-bridge to easily attach to the base layer metal (because they are made of the same material, Au), in practice, we found that air-bridges could easily break off from the base layer during the lift-off process. This was found to be caused by small pockets of residual resist which remains on the sample surface after development. We studied the sample surface roughness by AFM after trialling different development methods. Figure 4.6 shows a comparison of such a study. The images were taken over test patterns written on a layer of 950k PMMA. They reveal the variations in the developers' effectiveness at removing residual resists. It can be seen that the 3:7 water/IPA solution produces the best finished surface with the minimal residue, in agreement with other studies [102, 101]. These studies show that in addition to the superior finish, water/IPA is a high contrast developer, meaning that its rate of development varies strongly with the e-beam dose. For these reasons it was chosen for the air-bridge process.

Since MMA/MAA has much higher e-beam sensitivity than 950k and 100k PMMA, the middle layer is developed at a very high rate once the developer has dissolved the top layer of resist, making it difficult to accurately control the timing of the development progress. We can resolve this difficulty by lowering the development temperature, thereby limiting the rate of development. In practice, this was achieved by cooling the water/IPA solution in a glass bottle in a constant-temperature water bath at 5 °C.

4.4.2 Modelling and calibration of development time

In order to determine the pedestal and bridge dose given the choice of resists and developer, we built a numerical model with MATLAB to calculate the thickness of the resist structure as a function of development time and e-beam dose. The model assumes that the development rate of the resists is time-independent and is determined by the e-beam dose plus the proximity effect. It also assumes that the development proceeds in the direction normal to the resist surface at all times.

The development rate of each type of resist was determined by calibration measurements on test samples. Plain GaAs chips were coated with single layers of each type of resist, and

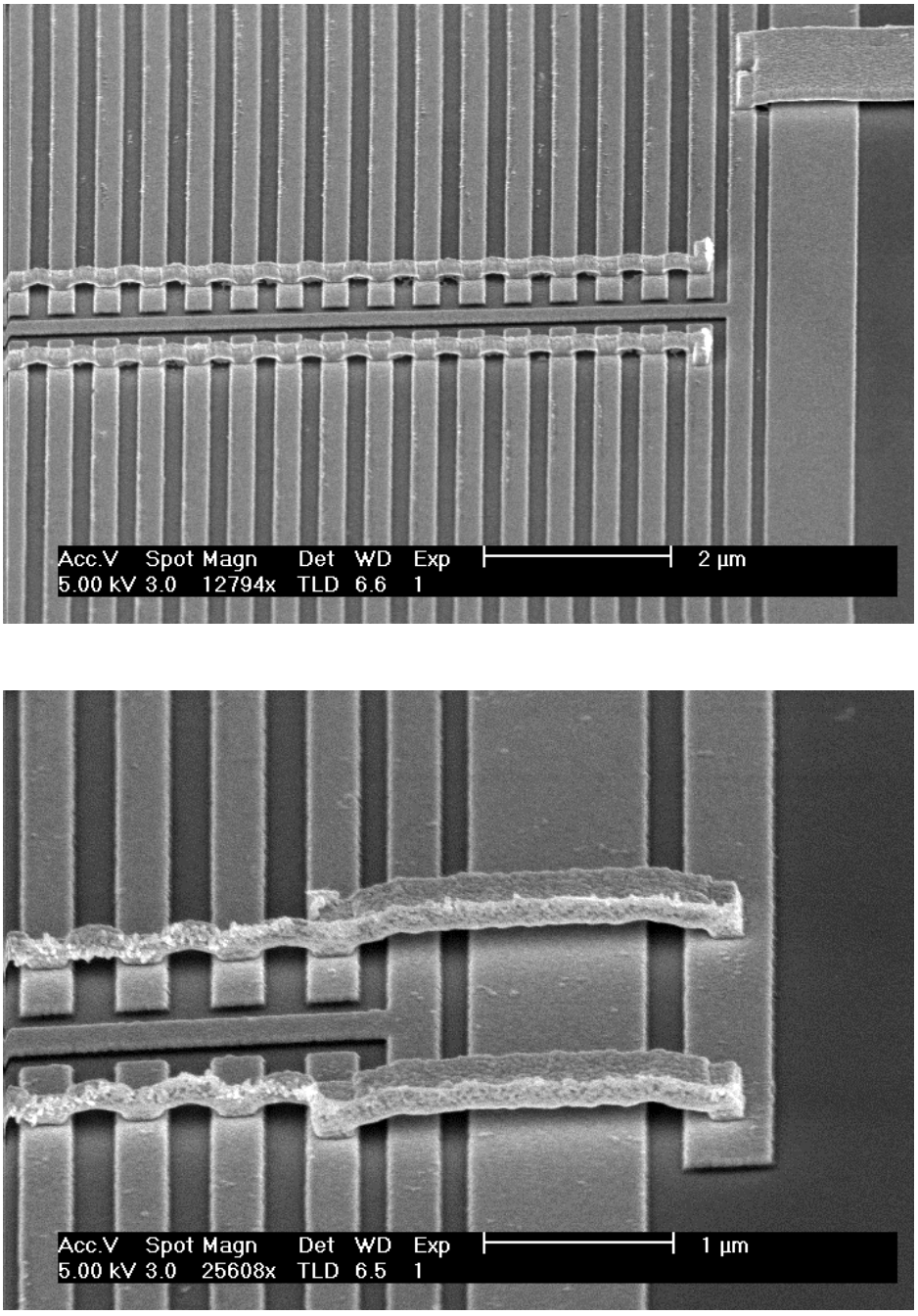


Figure 4.5 SEM images of air-bridges on working devices.

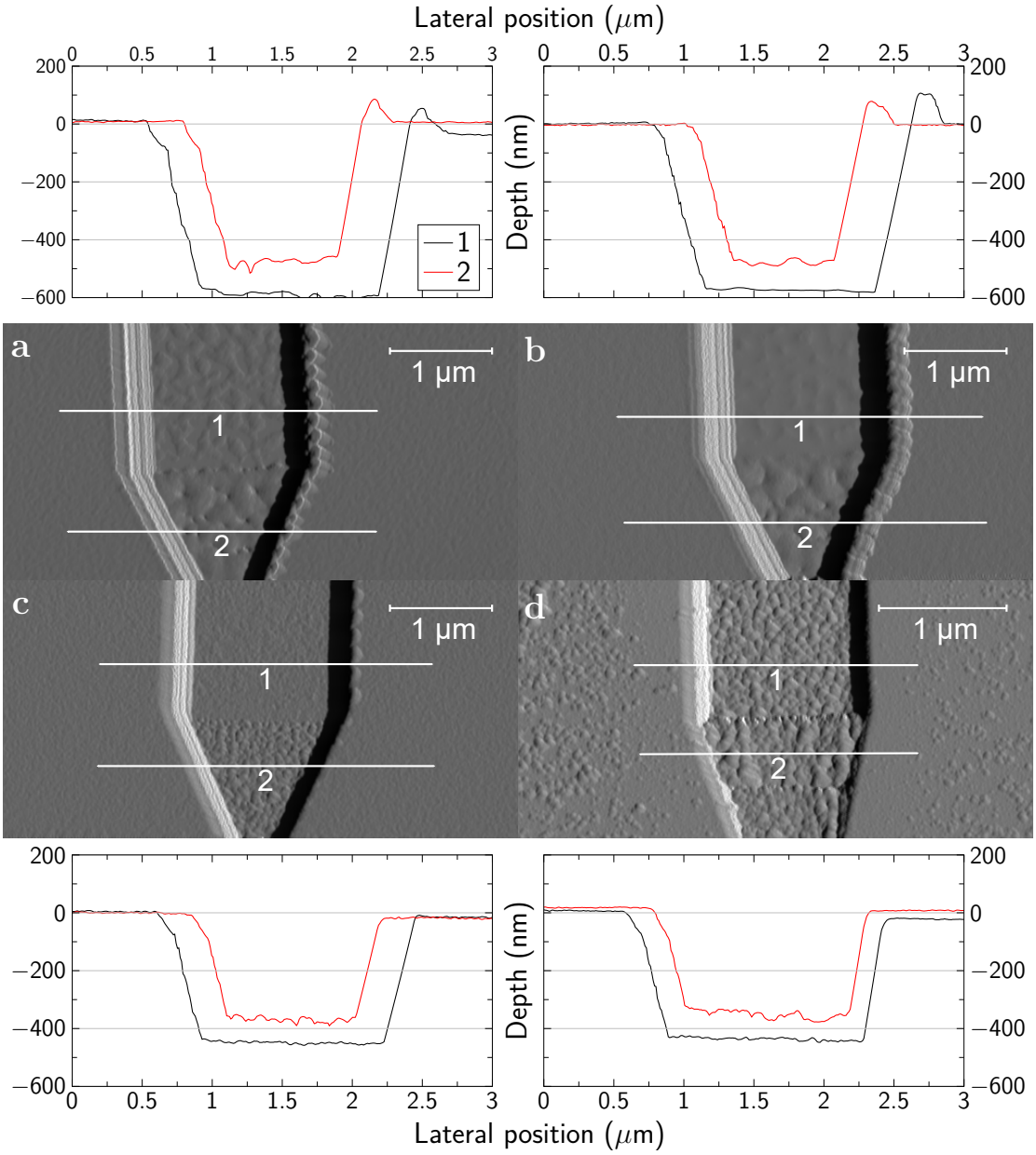


FIGURE 4.6 A comparison of AFM images of 950k PMMA after different development processes. The images are generated from the differential AFM data, which better demonstrate the surface roughness. The cross-sectional profiles attached to each image were measured across the horizontal lines.

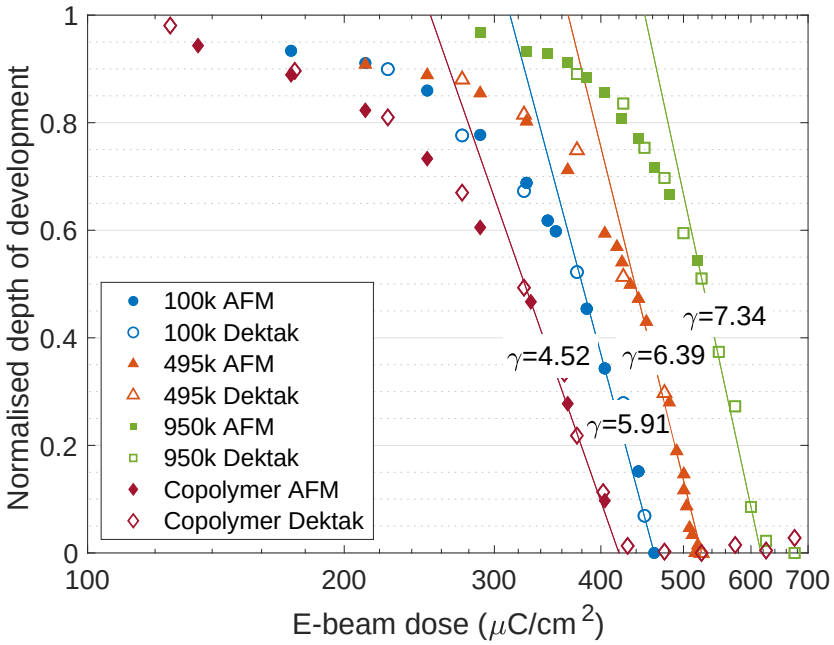


FIGURE 4.7 Normalised depth of developed resists plotted as functions of the e-beam dose after 30 s of development in 3:7 water/IPA at 5 °C. The Dektak data series were measured from large rectangular exposure patterns while the AFM series were measured from fine grating patterns with 50% coverage. It was found that in order to make the two series coincide, the dose of the AFM series needed to be scaled by a factor of 0.77 (see the discussion following Equation 4.9).

were exposed by a wide range of e-beam doses in 200 μm × 300 μm rectangular patterns and a grid of fine gratings of dimension 2 μm × 300 μm, separated by 2 μm gaps. The test samples were developed in 3:7 water-IPA solution at 5 °C with no ultrasound assist for 30 seconds. The depth of the rectangular patterns was measured with a Dektak surface profiler while the depth of the gratings was measured by AFM, as the stylus of the surface profiler was larger than the gratings. Figure 4.7 shows the results of the depth measurements. The developed depths of each type of resist were normalised to their total applied thicknesses: 146 nm for the 950k PMMA, 313 nm for the MMA/MAA co-polymer and 105 nm for the 100k PMMA, measured by ellipsometry. The e-beam doses are plotted on a logarithmic scale. The gradient γ of the linear portions of the plots at the high-dose end is known as the contrast of the resists. The values obtained in our measurement are similar to those reported in [74]. By applying a polynomial fit to each of these curves, we can estimate the development rate at arbitrary dose.

Proximity effects

During e-beam exposure, the high-energy electrons leaving the source (also called the primary electrons) can easily penetrate the resist. After entering the resist, a portion of them is reflected directly out of the substrate while losing little momentum. These are called back-scattered electrons and their effect of breaking down the chemical bonds of the resist is limited to within a small radius from the centre of the beam. For the rest of the electrons that enter the substrate, collision with the crystal lattice causes heavy scattering, resulting in significant loss of momentum and energy. Eventually some electrons re-emerge from the substrate and enter the resist a second time. These are known as secondary electrons. They are distributed over a much wide area around the original beam and cause further exposure to the resist. In simulating the e-beam exposure, one has to consider the effects of both the forward and back-scattered electrons in order to determine the total amount of dosage received at any point on the resist. Exposing a new feature close to existing patterns causes further exposure to an already-exposed pattern. This is known as the proximity effect. Commercial software packages (such as BEAMER developed by Genisys GmbH, which is used on the Vistec system) use Monte-Carlo simulation of the scattering process to determine the effects of different electrons and adjust the applied dose to match the intended dose. In the following paragraph, we present a simple model which treats the distribution of primary, back-scattered and secondary electrons as two Gaussian functions and was used in our numerical model.

Consider the process of patterning an array of linear features (Figure 4.8). The dose Q applied to a unit area of sample at position (x, y) due to an e-beam tracing a pattern nearby is given by the sum of the dose from exposing each pixel (i. e. a convolution) along the scanning path:

$$Q(\boldsymbol{r}) = \int_{x_0}^{x_1} \mathrm{d}x' \int_{y_0}^{y_1} \mathrm{d}y'\, d(x', y') q(\boldsymbol{r}), \tag{4.2}$$

where the primed coordinates denote the position of the beam centre. $d(x', y')$ is the dose function, which specifies the e-beam dose needed at (x', y'). It is a top-hat function inside the boundaries of the linear pattern (delimited by x_0, x_1, y_0, y_1). $q(\boldsymbol{r})$ is the point-spread function (PSF), which describes the broadening of the beam. $r = \sqrt{(x - x')^2 + (y - y')^2}$ is the distance between the point under consideration and the beam centre.

The double-Gaussian model treats the PSF as the sum of two Gaussian functions:

$$q(\boldsymbol{r}) = A \left[\exp\left(-\frac{r^2}{\alpha^2} \right) + \eta \exp\left(-\frac{r^2}{\beta^2} \right) \right], \tag{4.3}$$

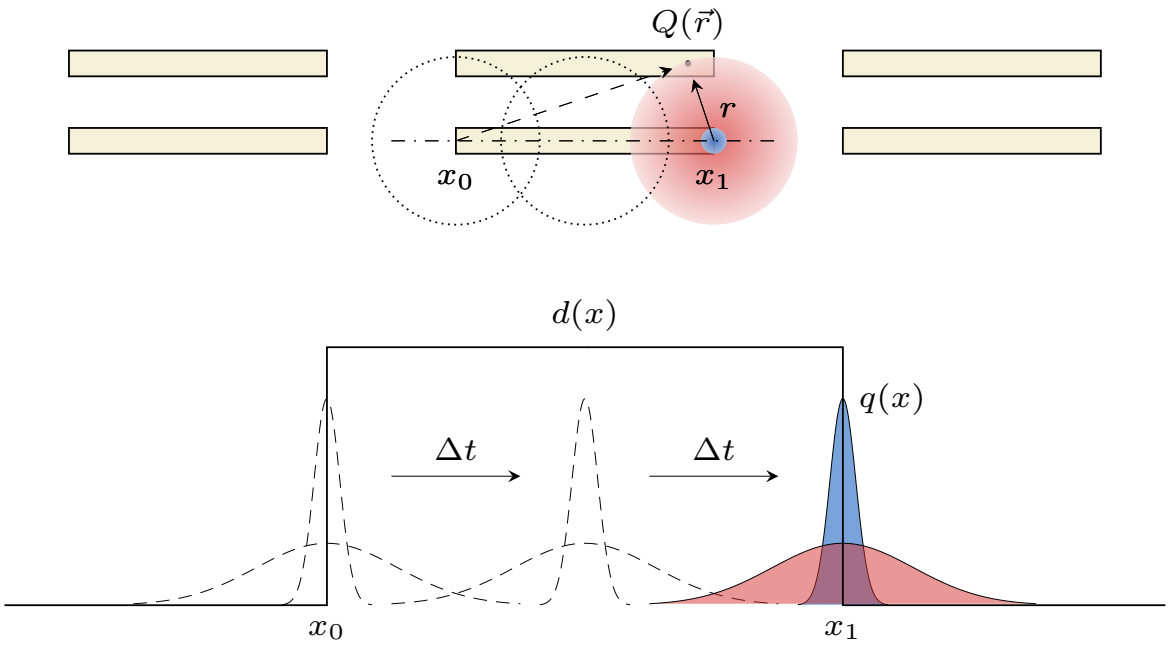


FIGURE 4.8 Double-Gaussian model of e-beam proximity effect. The yellow patterns in the top half of the plot represent the array to be exposed by e-beam. The lower half of the plot shows the dose function taken along the dash-dotted line in the x direction. The dose function $d(x, y)$ describes the desired dose distribution on the sample, which is a series of top-hat functions. The point-spread function $q(x, y)$ consists of two Gaussian terms, which describe the actual e-beam dose applied by the electron gun. Note the spread of the primary beam (blue) and the secondary beam (red) are not drawn to scale–they differ by four orders of magnitude. The e-beam dose (measured in $\mu C\,cm^{-2}$) received at any position on the sample is denoted by $Q(x, y)$, and is equal to the sum of the PSF as the beam is scanned along the array.

where α represents the broadening of the primary and back-scattered electrons, while β represents that of the back-scattered electrons. η is the ratio between the energy of the backward scattered and the forward electrons. The constant A can be determined by normalising the PSF with respect to the equivalent overall dose $\mathcal{D}$ (taking the definition of the dose to be that at the centre of the large-area shape):

$$\mathcal{D} = \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y'\, q(\boldsymbol{r}) = A(\alpha^2\pi + \eta\beta^2\pi) \quad \Rightarrow \quad A = \frac{\mathcal{D}}{\pi(\alpha^2 + \eta\beta^2)}, \tag{4.4}$$

where we have used the result $\int_{-\infty}^{\infty} \mathrm{d}x'\, e^{-(x-x')^2/\alpha^2} = \alpha\sqrt{\pi}$.

Substituting this into equation 4.2, we have the expression of the dose due to exposure of a single line:

$$Q(\boldsymbol{r}) = \int_{x_0}^{x_1} \mathrm{d}x' \int_{y_0}^{y_1} \mathrm{d}y' \, \frac{\mathcal{D}}{\pi(\alpha^2 + \eta\beta^2)} \left\{ \exp\left[-\frac{(r - r')^2}{\alpha^2} \right] + \eta \exp\left[-\frac{(r - r')^2}{\beta^2} \right] \right\}. \quad (4.5)$$

Since we have normalised the dose, $d(x, y)$ is equal to unity inside the pattern and zero elsewhere. Note that the values of α and β differ significantly: fitting the Monte-Carlo simulation of BEAMER for GaAs to a Gaussian function, we find $\alpha = 2.5\,\mathrm{nm}$ at an acceleration voltage of 100 kV, which is much smaller than the typical size of e-beam features (0.05 µm - 0.5 µm). On the other hand, $\beta = 11.37\,\mu\mathrm{m}$. The Gaussian integral over a finite limit can be evaluated numerically by invoking the error function in most numerical packages (MATLAB in our case):

$$\begin{aligned} &\int_{x_0}^{x_1} \mathrm{d}x' \int_{y_0}^{y_1} \mathrm{d}y' \, \frac{\mathcal{D}}{\pi(\alpha^2 + \eta\beta^2)} \exp\left[-\frac{(r - r')^2}{\alpha^2} \right] \\ &\quad = \pi\alpha \left[\mathrm{erf}\left(\frac{x - x_1}{\alpha} \right) - \mathrm{erf}\left(\frac{x - x_0}{\alpha} \right) \right] \cdot \beta \left[\mathrm{erf}\left(\frac{y - y_1}{\beta} \right) - \mathrm{erf}\left(\frac{y - y_0}{\beta} \right) \right] \end{aligned} \quad (4.6)$$

Alternatively, we can argue that since $\alpha \ll |x_1 - x_0|$, the majority of the sharp Gaussian function is within the finite limit. The first term of Equation 4.5 would remain numerically the same if we make the limits of the integral infinite:

$$\int_{x_0}^{x_1} \mathrm{d}x' \int_{y_0}^{y_1} \mathrm{d}y' \, \frac{\mathcal{D}}{\pi(\alpha^2 + \eta\beta^2)} e^{-(r-r')^2/\alpha^2} = \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y' \cdots = \frac{\mathcal{D}\alpha^2}{\alpha^2 + \eta\beta^2}. \quad (4.7)$$

On the other hand, since $\beta \gg |x_1 - x_0|$, the second term manifests itself as a constant background, which can be approximated by the average dose across the whole sample:

$$\int_{x_0}^{x_1} \mathrm{d}x' \int_{y_0}^{y_1} \mathrm{d}y' \, \frac{\mathcal{D}}{\pi(\alpha^2 + \eta\beta^2)} \eta e^{-(r-r')^2/\beta^2} = c \int_{-\infty}^{\infty} \mathrm{d}x' \int_{-\infty}^{\infty} \mathrm{d}y' \cdots = \frac{\mathcal{D}c\eta\beta^2}{\alpha^2 + \eta\beta^2} \quad (4.8)$$

where c is the coverage of the e-beam patterns, i. e. the ratio between the area of the e-beam pattern to the total sample area. The limits of the second integral are again replaced by infinities, because the sample is much larger than a single e-beam pattern. Collecting terms, we arrive at a relationship between $Q(\boldsymbol{r})$, which corresponds to the dose received by the resist,

and $\mathcal{D}$, which corresponds to the dose applied by the e-beam source:

$$Q(\boldsymbol{r}) = \frac{\alpha^2 + c\eta\beta^2}{\alpha^2 + \eta\beta^2}\mathcal{D} = \frac{1 + c\eta'}{1 + \eta'}\mathcal{D}, \tag{4.9}$$

where $\eta' = \eta\beta^2/\alpha^2$ is a factor which solely depends on the beam scatterings. As was pointed out in the caption of Figure 4.7, for GaAs substrate and a coverage of $c = 1/2$, the dose given by the proximity correction software was different by a factor of 0.77 to the requested dose. Applying this to the above formula results in $1 + c\eta'/1 + \eta' = 0.77$, which corresponds a solution of $\eta' = 0.85$.

Equation 4.9 can be generalised to cases where more than one dose is applied:

$$Q_i(\boldsymbol{r}) = \frac{\mathcal{D}_i}{1 + \eta'} + \sum_i \frac{\eta'}{1 + \eta'} c_i \mathcal{D}_i \implies$$

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \frac{1}{1 + \eta'} \left[\begin{pmatrix} \mathcal{D}_1 \\ \mathcal{D}_2 \end{pmatrix} + \eta' \begin{pmatrix} c_1 & c_2 \\ c_1 & c_2 \end{pmatrix} \begin{pmatrix} \mathcal{D}_1 \\ \mathcal{D}_2 \end{pmatrix} \right], \tag{4.10}$$

where the index i in the first part labels the different doses. In the second part of the expression, we have rewritten the general expression for a two-dose case (such as the air-bridge process). Equation 4.10 shows that the proximity effect can be considered as an operator which transforms the applied dose $\mathcal{D}_i$ into the received dose Q. By inverting this operator (usually by computer), we may calculate the dose function required to achieve the desired actual exposure, with the limitation that the dose values cannot be negative.

Figure 4.9 shows the results of the numerical model. We found that despite the model predicting that a 60 s-long development will eliminate all three layers of resist, in reality, samples needed to be developed for at least that long to clear the pedestals so as to produce strong metal to metal contact between the air-bridges and the base gates. One explanation for this discrepancy is that the development rate used in the calculation was based on calibration tests which ran for 30 s instead of 60 s. Our assumption of the rate of development being time-independent is likely to be erroneous, because the development process should gradually slow down: as the developer dissolves more resist, it becomes saturated with the solute so the rate of dissolution would decrease. Further calibration at 60 s was not performed, as a suitable time had been found by trial. However, the numerical model remains a valuable tool should a process with different resist thickness be attempted in the future. It is worth noting that the numerical model does correctly predict the shape of under-developed resist, as can be seen from the SEM image of Figure 4.10.

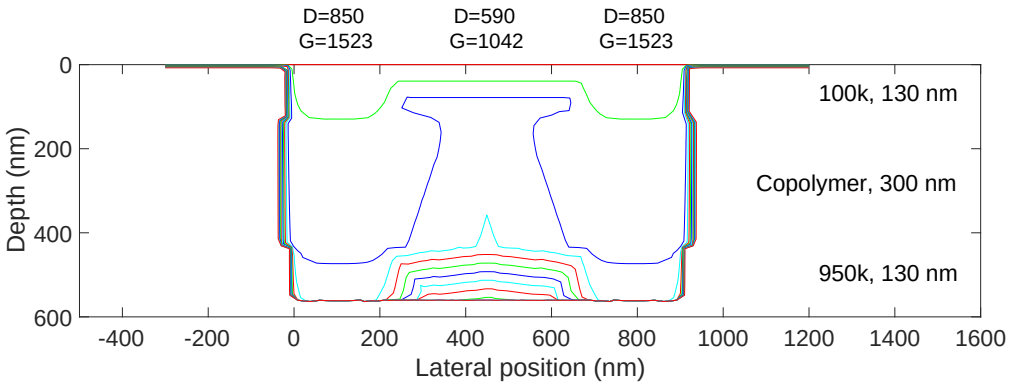


FIGURE 4.9 The evolution in time of a stack of triple-layer resist as calculated by the numerical model. The development time was 60 s. Adjacent profiles are 5 s apart. The labels at the top indicates the e-beam doses applied to the corresponding regions on the sample. The values indicated by D are the dose required at each position and the values indicated by G are the corresponding proximity-corrected doses. The figures on the right show the thickness of each layer of resist.

Sample packaging

The last step of fabrication involves bonding the device to a sample package. Individual devices were scribed off the processed chip and cleaned in acetone before being glued to a 20-pin ceramic lead-less chip carrier (LCC) package with GE varnish. Connections to the LCC pins were made with gold wires on a bonding machine. Care must be taken during this process to eliminate electrostatic discharge (ESD), which can easily damage the intricate e-beam patterns in the narrow parts of the structure. To minimise such damages, the following measures were taken: 1. Running air ionisers in the vicinity of the bonding machine. 2. Grounding the LCC package and tools such as tweezers to the body of the bonding machine. 3. Always bond the contacts furthest away from the previous connection, so as to allow time for any charge to dissipate from the sample. 4. Using a wedge bonder. It is significantly less likely to cause ESD damage compared with a ball bonder, which operates a 'sparking' action that uses a high voltage discharge to create a spherical tip on the open end of the bond wire. This action invariably creates an ion-rich environment around the sample, which increases the chance of ESD. Wedge bonding on the other hand, relies purely on ultrasonic vibrations to fuse the bond wires with the pins on the LCC package, hence is much less likely to result in ESD damage.

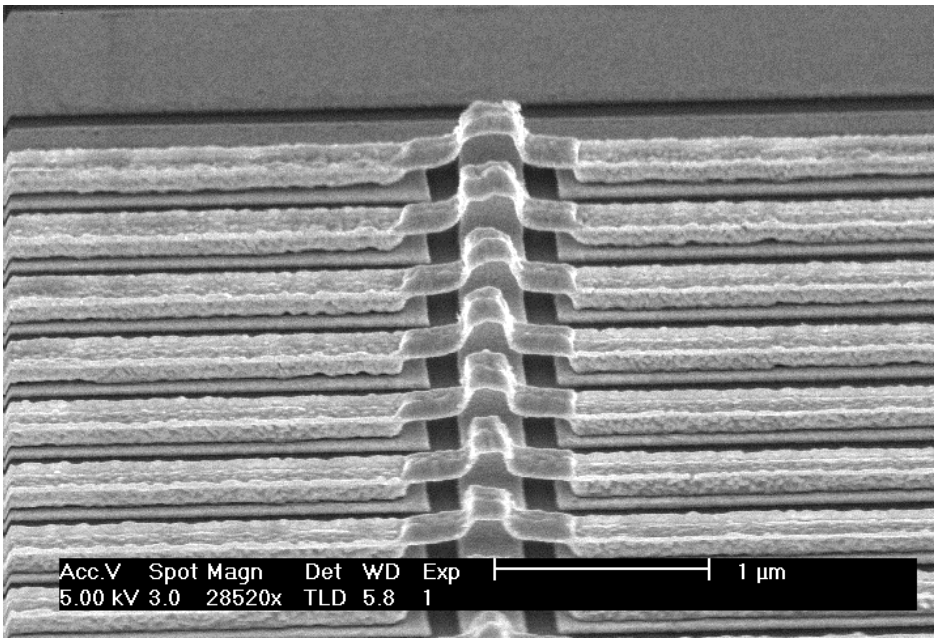


Figure 4.10 SEM image showing air-bridges formed from under-developed resist.

4.5 Summary

In this chapter, we have reviewed the fabrication details of our experimental device by following the process from wafer cutting to packaging. In the author's experience, a complete fabrication cycle can take up to several weeks, therefore it is very important to be aware of some of the common pitfalls of fabrication. For example, it is easy to miss a seemingly trivial process such as the 10-second acid dip for oxide removal, and the consequence of such a mistake would not be realised until the much later stage of wire bonding, where the bond pads may struggle to bond because of poor surface adhesion. Apart from rigorously following a fabrication recipe, the author would like to suggest processing plain wafers as test devices together with real samples as a way of quality control. Test samples would prove useful, for example, in the hypothetical scenario of oxide removal being missed: while we cannot start bonding a real sample after the optical stage of processing, a dummy sample can be used immediately to test the quality of bond pads, and reveal the problem of poor adhesion before we continue to the EBL stages of the process.

Chapter 5

Measurement equipment and techniques

5.1 Cryogenic systems

The energy associated with the thermal motion of a microscopic particle is $k_B T = 25.7$ meV at room temperature ($T = 298$ K). It is therefore difficult to measure quantum-mechanical properties of electrons under this condition, as their typical energy scale is only a few meV. For example, the Fermi energy with one electron subband occupied in our sample is measured to be about 4 meV. To achieve a good signal-to-noise ratio, measurements need to be carried out at temperatures significantly lower than 300 K with the use of cryogenic systems. The following sections outline the various cryogenic systems employed in this study.

5.1.1 ^{4}He dipping dewar

The simplest low temperature measurement is carried out by immersing the device directly in liquid helium-4, which has a boiling point of 4.2 K at atmospheric pressure. The measurement setup involves a long sample probe which has a 20-pin Charntek LCC package holder on one end and a breakout box on the other. The package holder is a socket with spring-loaded metal pins, into which an LCC package can be inserted. When the sample is loaded, each pin is pressed onto a matching contact on the LCC package, securing it as well as making electrical contact. The pins are connected to a loom of copper (or constantan) wires which are laid inside the cavity of the hollow probe, and are attached to the breakout box at the other end. There are 20 BNC connectors on the box, each connected to a pin of the package by the copper wires. The input and output terminals of different instruments can be connected to the sample via these BNC connectors using coaxial cables.

During measurement, the sample is loaded onto the probe and lowered into a liquid-helium dewar. As long as it is maintained under the liquid-helium surface, its temperature

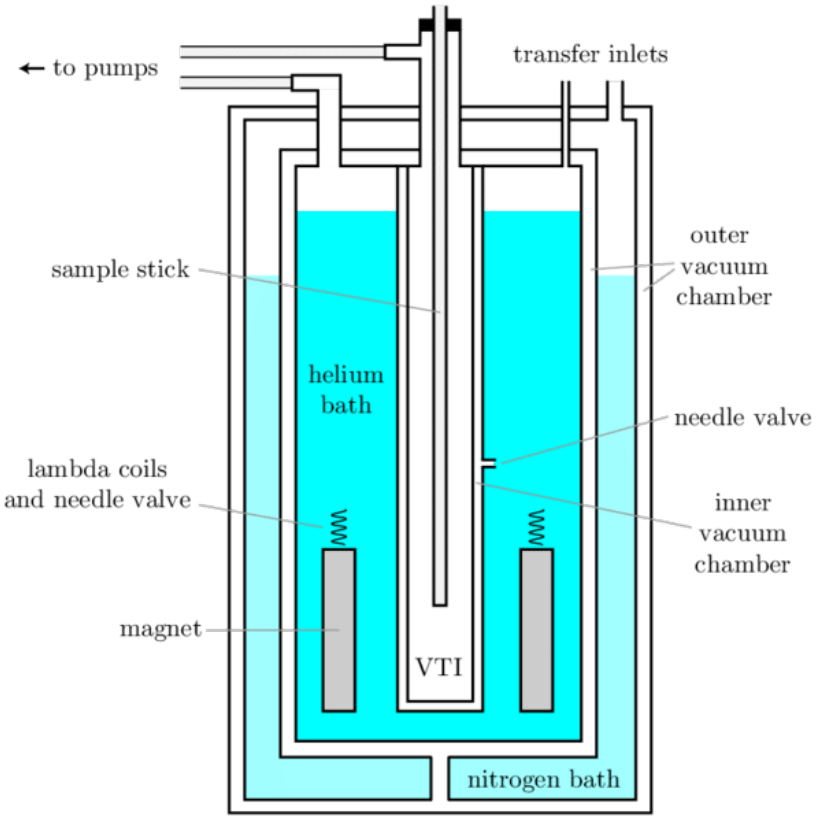


FIGURE 5.1 Schematics of a lambda-point cryostat. Image from [38].

remains at 4.2 K. Due to its simplicity, the dipping dewar was used to perform preliminary checks on a device immediately after fabrication. The conductance of a sample was measured with a source-measure-unit (SMU) to check that the ohmic contacts conduct properly. Bias voltage was applied to the various top gates to check they function as designed. Samples that pass the checks were marked for detailed measurements in advanced systems later.

5.1.2 Pumped ^{4}He cryostat

The pumped ^{4}He cryostat is a system where the sample is cooled in a helium vessel that is continuously evacuated by a vacuum pump. Since the boiling point of helium is lowered by reducing the vapour pressure [7], the system operates below 4.2 K. As shown in Figure 5.1, the cryostat consists of a vacuum-shielded sample chamber and is called a VTI (variable temperature insert) and isolated inside a shielded liquid-helium bath. The shielding consists of two evacuated spaces: the outer vacuum chambers (OVC), and a liquid-nitrogen bath between them. ^{4}He has a very low specific latent heat of vaporisation of 20 kJ kg^{-1}, about 10% to that of nitrogen and 1% of water. The double shields reduce liquid helium loss by reducing

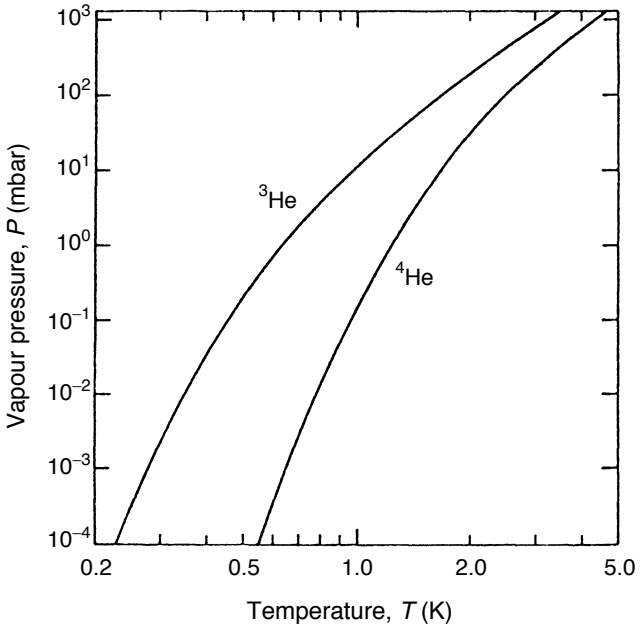


FIGURE 5.2 Vapour pressure of ^{3}He and ^{4}He as functions of temperature. Image from [72].

the thermal radiation from its surroundings–according to Stefan's law, the power of thermal radiation is proportional to T^4, and liquid nitrogen has boiling point 77 K, significantly lower than 298 K. The sample space is connected to the helium bath by a needle valve whose function is to admit a regulated flow of liquid helium from the bath to the sample space. During operation, the sample is lowered into the sample space on a probe, and this space is continuously evacuated by a rotary pump. A small stream of liquid helium is admitted until the temperature has dropped to 10 K, at which point the needle valve is fine-tuned so that the sample-space pressure is 3 to 5 mbar. A stable state is reached in approximately 30 minutes, with an operation temperature between 1.2 to 1.5 K.

The ^{4}He cryostat used in this work was a lambda-point cryostat manufactured by Oxford Instruments. The system contains a superconducting magnet at the bottom of the helium reservoir, which can be used to provide a field along the length of the VTI. The lambda point operation refers to the action of pumping on the helium bath around the magnet to cool it to 2.17 K, which boosts the performance of the magnet at the expense of increased helium consumption. This mode was not used in our study, and the system was used to perform measurements with magnetic field up to 10 T.

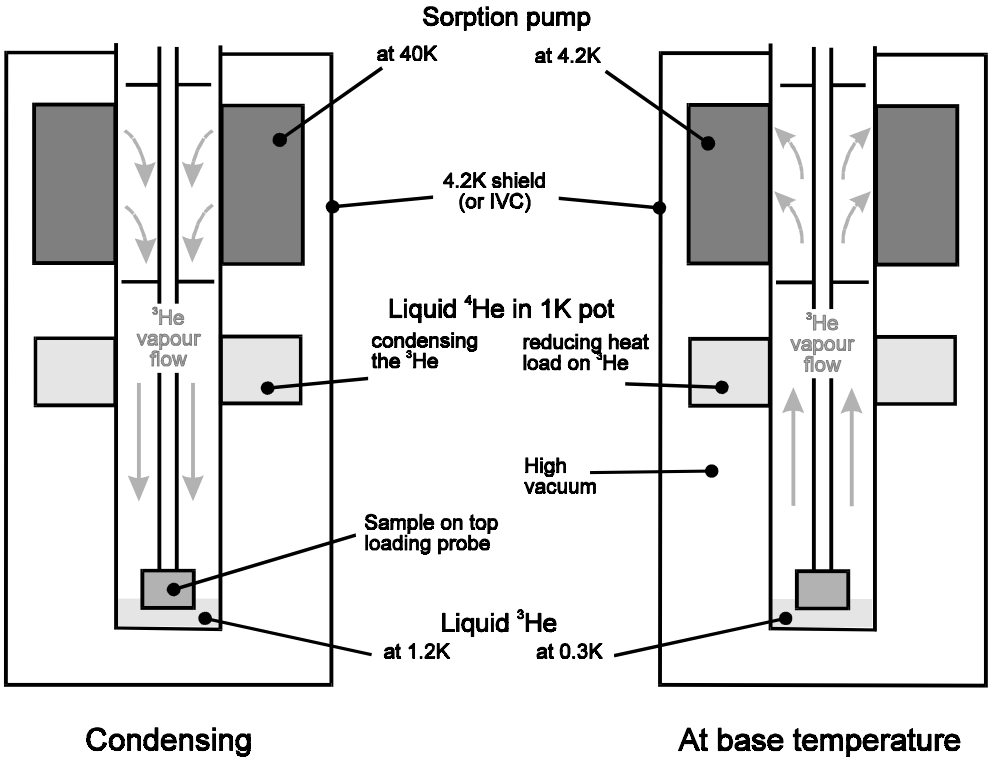


FIGURE 5.3 Operation of a sorption-pumped ^{3}He cryostat. Image from [7].

5.1.3 Sorption-pumped ^{3}He cryostat

Temperatures lower than that of the pumped ^{4}He cryostat can be achieved by substituting the ^{4}He cryogen with its lighter isotope ^{3}He (Figure 5.2), which has lower boiling point at the same vapour pressure. Due to its scarcity (hence high cost), ^{3}He must be used in a closed-cycle system instead of rotary-pumped. The sorption-pumped ^{3}He cryostat is a top-loading, vacuum-sealed unit which replaces the VTI on a pumped ^{4}He system. As shown in Figure 5.3, it consists of a condenser and a sorption pump and contains the ^{3}He. The condenser is a heat exchanger located inside the 1K pot, a chamber which is fed by the main ^{4}He bath via a needle valve. The 1K pot is cooled by rotary-pumped ^{4}He. Its function is to condense gaseous ^{3}He inside the condenser into liquid. Liquefied ^{3}He collects at the bottom of the system and once the majority of the gas is liquefied, the sorption pump located at the top of the insert is turned on. The sorption pump (or sorb) is made of porous material that adsorbs ^{3}He atoms when it is at temperatures below 35 K. The vapour pressure at the sorb is dependent on its temperature. By cooling the sorption pump with ^{4}He at 4.2 K via heat exchanger, the vapour pressure is minimised and then a base temperature of 300 mK can be achieved at the bottom of the insert, which is where the sample is located.

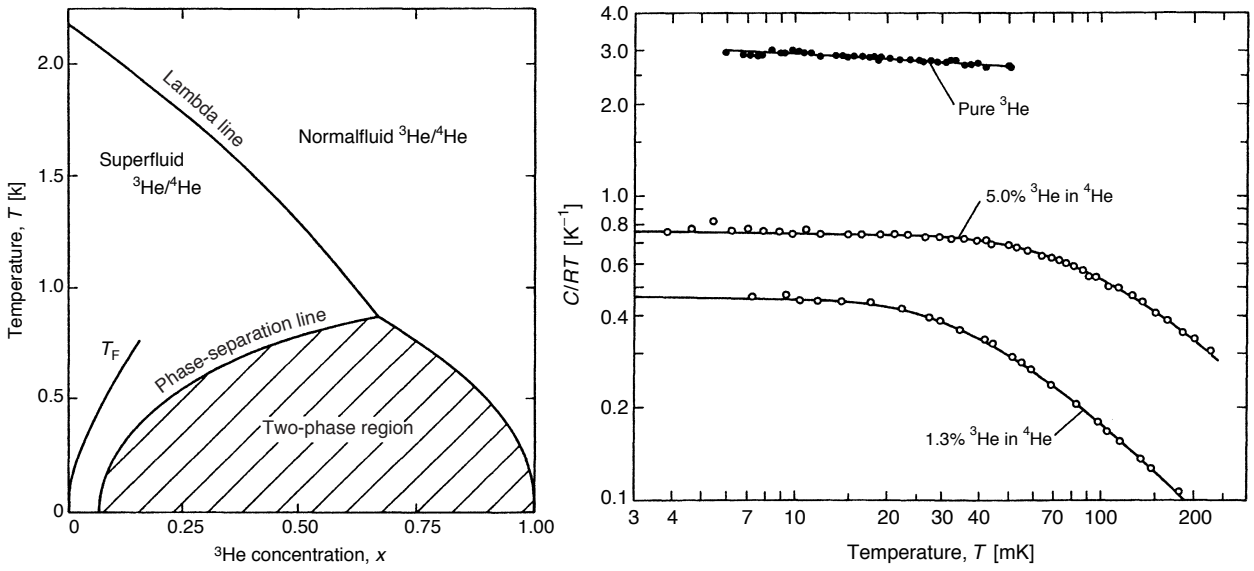


FIGURE 5.4 Left: Phase diagram of ^{3}He/^{4}He mixture at saturated vapour pressure. The shaded region marks the concentration at which the mixture is forbidden to exist. Right: Specific heat capacity of ^{3}He/^{4}He mixture at different concentration of ^{3}He. Image from [72].

The sorption pump cryostat is a single-shot system, which has limited operation duration. The finite amount of ^{3}He contained in the insert will eventually be captured by the sorb. The cooling stops once the liquid ^{3}He is exhausted and the sorption pump needs to be heated to above 35 K in order to release the trapped ^{3}He. The operation of the cryostat alternates between condensation and cooling, where the former requires about 1 hour and the latter can be maintained for about 12 or more hours. The ^{3}He cryostat used in this work is manufactured by Oxford Instruments. A superconducting magnet is located outside the sample space in the main helium bath and is capable of providing magnetic fields up to 14 T.

5.1.4 ^{3}He/^{4}He dilution refrigerator

The most sophisticated cryostat employed in this work was a ^{3}He/^{4}He dilution refrigerator, a system which utilises a continuous cycle of mixing and separation of two helium isotopes for cooling. At 2.17 K (referred to as the Lambda point), ^{4}He undergoes a second-order phase transition where its specific heat capacity C at saturated vapour pressure diverges (the phase diagram of helium resembles the letter λ around the singularity, hence the name lambda point). As temperature is lowered below the lambda point, an increasing fraction of ^{4}He becomes a superfluid (known as ^{4}He–II) with zero viscosity and entropy [17]. As for the fermionic ^{3}He, the transition to the superfluid phase requires the formation of bosonic Cooper pairs of atoms, which occurs at the much lower critical point of 2.49 mK–a tem-

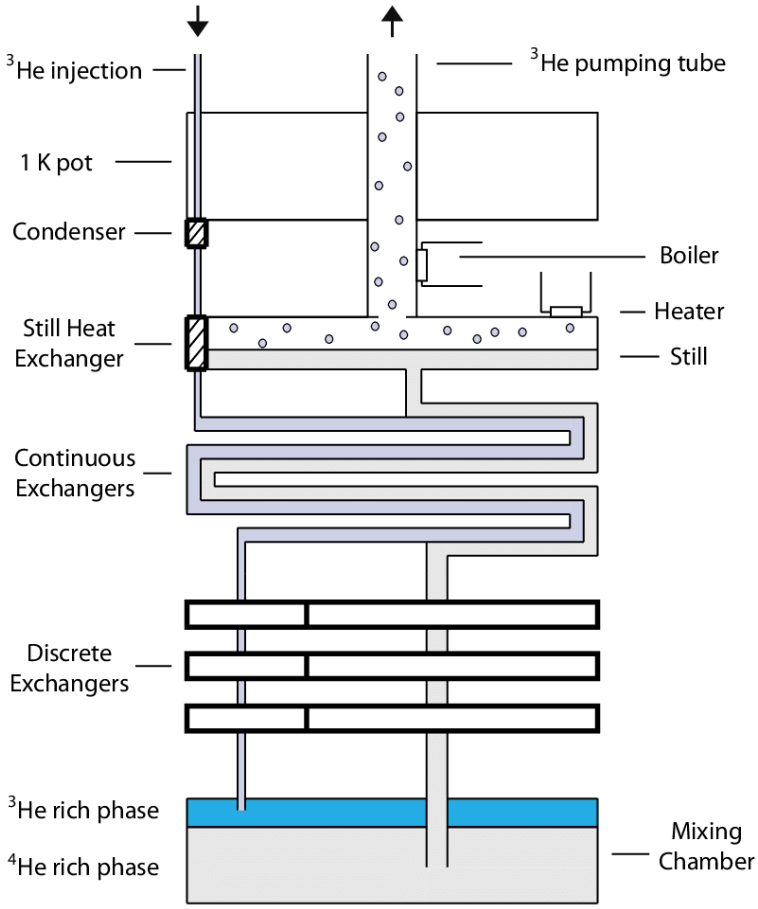


FIGURE 5.5 Schematics of a ^{3}He/^{4}He dilution refrigerator. Image from [6]

perature unachievable in the fridge. As the temperature drops below 2.17 K, the increasing fraction of ^{4}He–II separates from the normal mixture of ^{3}He and ^{4}He in the fridge. As shown in Figure 5.4, the ^{3}He/^{4}He mixture will spontaneously separate into two phases in the mixing chamber at $T < 870$ mK: one of low concentration ^{3}He in ^{4}He–II (the dilute phase) and another of almost pure ^{3}He in ^{4}He–I (the concentrated phase). The cooling power comes from the flow of ^{3}He atoms from the concentrated phase to the dilute phase: the energy removed by the process is given by the enthalpy change of the mixture, which is equal to $\Delta H = \int \mathrm{d}T\, C(T)$. As shown in Figure 5.4, the heat capacity of the concentrated phase is higher than that of the dilute phase, hence the mixing process is endothermic with $\Delta H > 0$. In other words, ^{3}He atoms take energy out of the concentrated phase as they evaporate into the dilute phase.

Figure 5.5 illustrates the main components of a dilution refrigerator. As for the VTI of a pumped ^{4}He cryostat, it is housed in a double-shielded helium reservoir which is not shown in the figure. Before entering the fridge, gaseous He (usually ^{3}He) at room temperature is

filtered by cold traps at liquid nitrogen (77 K) and helium (4.2 K) temperatures to remove any other gases, water or oil. The ^{3}He gas then enters the cryostat and goes into the condenser, which has a high-impedance capillary at its exit. The impedance ensures that the ^{3}He stays long enough in the condenser to be liquefied by the cooling power the 1K-pot. Liquid ^{3}He flows towards the mixing chamber at the bottom of the fridge, and is cooled further in a series of heat exchangers by the still (500 mK) and the cold dilute-phase mixture leaving the mixing chamber. The mixing chamber contains the phase-separated mixture and is the coldest part of the fridge. The sample holder is attached to the bottom of the mixing chamber on a 'tool' that goes inside the magnet. The nominal operating temperature of the fridge used in our experiment was 10 mK, although the actual temperature achieved was only 50 to 60 mK, as was measured by the thermometer in the fridge. The continuous flow of ^{3}He is achieved with the still, which is connected to the bottom of the mixing chamber by a pick-up tube and is evacuated by a diffusion pump. Since ^{3}He has far higher partial pressure than ^{4}He at the still temperature, the vapour in the still is pure ^{3}He. A heating resistor attached to the still can be turned on to facilitate the evaporation. The pumping of vapour creates an osmotic pressure, which drives the flow of ^{3}He from the dilute mixture to vapour in the still, and from the concentrated phase to the dilute-phase liquid in the mixing chamber. The ^{3}He vapour displaced by the pump returns to the room-temperature inlet, thus completing the cycle.

5.2 Wafer characterisation

Transport characteristics such as carrier density and mobility need to be determined before the wafer material is processed into experimental devices. The characterisation was based on a 4-terminal measurement of the quantum-Hall and the Shubnikov-de-Haas effects, whose theoretical and experimental details are outlined in the following sections.

Drude model and the classical Hall effect

The classical Hall effect can be observed when the magnetic field is small and perpendicular to the sample plane. It is explained by the Drude model, which applies the classical equation of motion to an electron under the effects of an electric field, a magnetic field and scattering:

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -e\boldsymbol{E} - e\boldsymbol{v}\times\boldsymbol{B} - \frac{m\boldsymbol{v}}{\tau}. \tag{5.1}$$

In steady state ($\mathrm{d}\boldsymbol{v}/\mathrm{d}t = 0$) the equation of motion becomes

$$\frac{e\tau}{m}\boldsymbol{v} \times \boldsymbol{B} + \boldsymbol{v} = -\frac{e\tau}{m}\boldsymbol{E}. \tag{5.2}$$

The current density through the Hall bar is $\boldsymbol{J} = -ne\boldsymbol{v}$, where n is the electron density, e the electronic charge, and $\boldsymbol{v}$ the drift velocity. In terms of $\boldsymbol{J}$, the solution to the steady-state equation above can be written as

$$\begin{pmatrix} 1 & \omega_\mathrm{c}\tau \\ -\omega_\mathrm{c}\tau & 1 \end{pmatrix} \boldsymbol{J} = \frac{ne^2\tau}{m}\boldsymbol{E}, \tag{5.3}$$

where $\omega_\mathrm{c} = eB/m$ is the cyclotron frequency. The matrix implies that, given a current in the x direction, there is an electric field in the y direction. This is the classical Hall effect.

Equation 5.3 has the form $\boldsymbol{E} = \boldsymbol{\rho}\boldsymbol{J}$, where $\boldsymbol{\rho}$ is the resistivity tensor. It can be written as

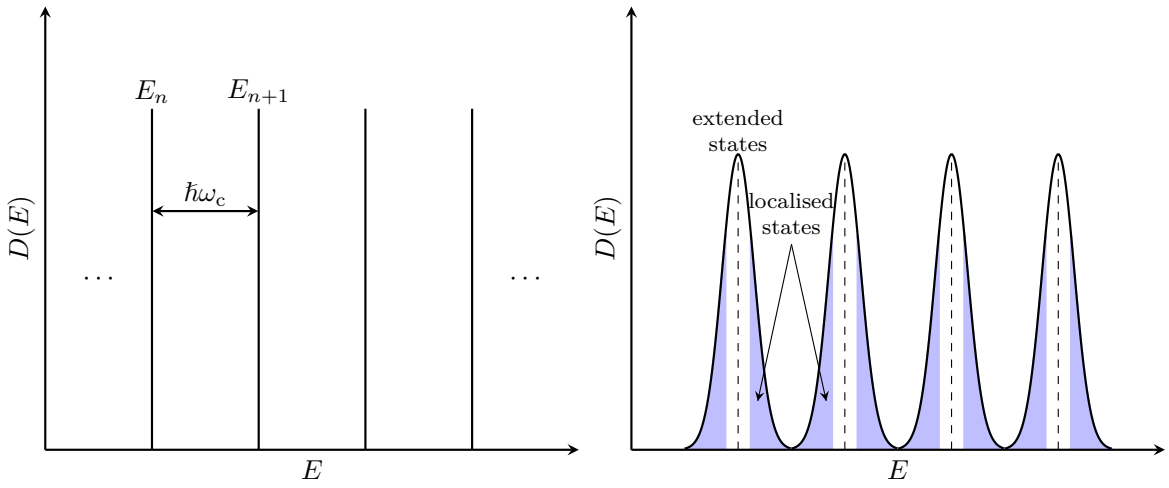
$$\boldsymbol{\rho} = \frac{1}{\sigma_\mathrm{DC}} \begin{pmatrix} 1 & \omega_\mathrm{c}\tau \\ -\omega_\mathrm{c}\tau & 1 \end{pmatrix}, \tag{5.4}$$

where $\sigma_\mathrm{DC} = ne^2\tau/m$ is the conductivity of the Hall bar in the absence of magnetic field. Since there is no transverse current flow in a Hall bar in the steady state ($J_y = 0$), we deduce from Equation 5.3 that $E_x = \rho_{xx}J_x$ and $E_y = -\rho_{xy}J_x$. Upon rearrangement, we have

$$\begin{aligned} \rho_{xx} &= \frac{E_x}{J_x} = \frac{V_x/L}{I_x/W} = \frac{V_x}{I_x}\frac{W}{L} = \frac{1}{\sigma_\mathrm{DC}} = \frac{m}{ne^2\tau}, \\ \rho_{xy} &= -\frac{E_y}{J_x} = -\frac{V_y/W}{I_x/W} = \frac{V_y}{I_x} = \frac{\omega_\mathrm{B}\tau}{\sigma_\mathrm{DC}} = \frac{B}{ne}, \end{aligned} \tag{5.5}$$

where L, W are the length and width of the sample under consideration.

Two important properties about the above results are: 1. The longitudinal resistivity ρ_{xx} is constant, while the transverse resistivity ρ_{xy} is linearly dependent on the magnetic field. The proportionality constant of ρ_{xy} to B can be used to determine the electron density; 2. In order to calculate the longitudinal resistivity, one needs the geometric parameters of the Hall bar, specifically the width-to-length ratio W/L, while the transverse resistivity is independent of any geometrical factors.



(a) Landau levels without scattering effects (b) Landau levels with scattering effects

FIGURE 5.6 The electron density of states of Landau levels. Left: without scattering, the electron states are concentrated in Landau levels with energy eigenvalues $E_n = \hbar\omega_B(n+1/2)$. Right: with scattering, the Landau levels are broadened so that the electron states have access to a spectrum of energies. Only those states with energies around a narrow band of the original Landau levels retain the plane-wave longitudinal wavefunction, which extend across the sample and contribute to conductance. The other states (shaded blue) orbit around impurities (also called localised states), which do not contribute to conductance.

Landau levels

In reality the Drude model result only applies at weak magnetic fields. When $\boldsymbol{B} > 0.5\,\mathrm{T}$, ρ_{xy} starts to increase in steps and remains on plateaux for extended ranges of $\boldsymbol{B}$; ρ_{xx} drops to 0 while ρ_{xy} is on a plateau and has spikes coinciding with the rising steps of ρ_{xy}. The behaviour of ρ_{xy} known as the integer quantum-Hall effect, and that of ρ_{xx} is known as the Shubnikov–de Haas effect.

The quantum-Hall effect is related to the formation of Landau levels. Under a magnetic field $\boldsymbol{B}$, the free-electron Hamiltonian acquires an additional momentum term:

$$H = \frac{1}{2m}(\boldsymbol{p} + e\boldsymbol{A})^2, \tag{5.6}$$

where $\boldsymbol{p}$ is the canonical momentum and $\boldsymbol{A}$ the magnetic vector potential for which $\nabla \times \boldsymbol{A} = \boldsymbol{B}$. For a rectangular sample geometry, one can use the Landau gauge $\boldsymbol{A} = yB\hat{\boldsymbol{x}}$, so that the Hamiltonian becomes:

$$H = \frac{1}{2m}\left[(p_x - eBy)^2 + p_y^2\right], \tag{5.7}$$

which is independent of the position operator x. This implies that the Hamiltonian commutes with momentum p_x. As a result, the wavefunction can be written as $\psi(x, y) = \mathrm{e}^{ik_x x} f(y)$.

Substituting this wavefunction into the Hamiltonian, we have:

$$H = \frac{1}{2m}\left[(\hbar k_x - eBy)^2 + p_y^2\right] = \frac{1}{2}m\omega_c^2\left(y - k_x l_c^2\right)^2 + \frac{p_y^2}{2m}, \tag{5.8}$$

which is a harmonic-oscillator-type Hamiltonian centred at $y = k_x l_B^2$, where $l_B = \sqrt{\hbar/eB}$ is called the magnetic length, and frequency $\omega_c = eB/m$ is the cyclotron frequency. The formula implies that the electronic states in a magnetic field are quantised into the transverse modes (y direction), with energy eigenvalues $E_n = \left(n + 1/2\right)\hbar\omega_c$, also referred to as the Landau levels (Figure 5.6(a)).

Conduction by edge states

Halperin [37] described the quantum Hall effect in terms of electron states that extend across the sample, along the edges of the conducting material. A simplified understanding of this can be gained by the following semi-classical consideration: the electron Hamiltonian in a Hall measurement is $H = (\boldsymbol{p} + e\boldsymbol{A})^2/2m + U(x, y)$, where $U(x, y)$ represents the confinement potential. We expect $U(x, y)$ to rises up sharply as y approaches the edges of the sample. Since all real materials invariably contain impurities, we expect $U(x, y)$ to fluctuate randomly in the bulk material. The confinement potential profile in the transverse direction, $U(y)$, is shown in Figure. 5.7(a). The group velocity of an electron state is proportional to the slope of its dispersion. In the longitudinal direction, it is given by

$$\hbar v_x = \frac{\partial E(n,k)}{\partial k_x} = \frac{\partial E(n,k)}{\partial y}\frac{\partial y}{\partial k_x} = \frac{1}{eB}\frac{\partial E(n,k)}{\partial y}, \tag{5.9}$$

where we have used $y = k_x l_B^2$ (the centre of the Landau levels) to calculate $\partial y/\partial k_x$. Since E is the sum of the kinetic and potential energies, its slope is directly related to the derivative $\partial U/\partial y$. As transport is determined by electrons near the Fermi energy, we consider the states at the intersection between the Landau levels and the Fermi level. From the slope of the potential depicted in Figure. 5.7(a), we see that electron states have velocities in the same direction on the same edge of the sample (and the directions are opposite on opposing edges). In contrast, states in the bulk of the sample do not have such uniformity in velocity. Following detailed discussions, for example [89], it can be shown that the drift velocity is perpendicular to $\nabla U(x, y)$. This suggests that states propagate along the equipotential lines. In the bulk material, the equipotentials can easily form closed paths around the peaks and troughs of the potential landscape, which are centred around impurities. States that drift along such closed equipotentials are localised and therefore do not contribute to conduction. On the contrary,

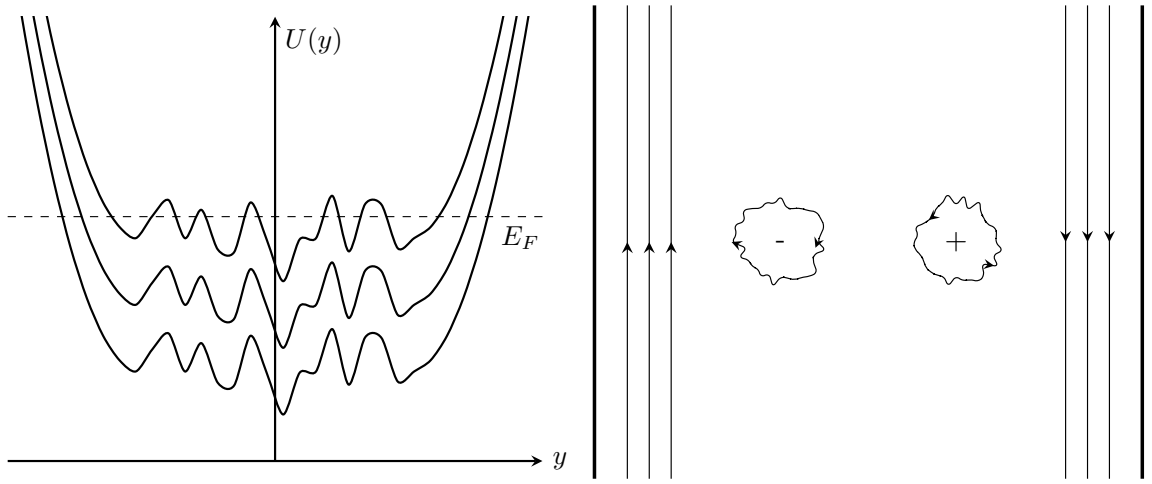


FIGURE 5.7 (a) Plot showing confining potential of three Landau levels with random perturbations in the bulk material due to impurity scattering. (b) Drift velocity of electron states in the sample follow confinement equipotentials. The '-' and '+' symbols represent a trough and a peak in the potential landscape, respectively. The plot shows that states in the bulk material are bound in localised orbits centred around the impurities, and states which extends across the whole sample are located at the edge of the sample. From the gradient of the potential, one can see that the states around the troughs move clockwise and the states around the peaks move anti-clockwise.

states that extend across the whole sample contribute to conduction. They are mostly found along the edge of the sample, because as discussed earlier, states along the edge of the sample all travel in the same direction (Figure 5.7(b)). The formation of edge channels happen at large $\boldsymbol{B}$: in the semi-classical picture, they are equivalent to the skipping orbits which form at high fields. For detailed quantum mechanical discussion on their formation, see Chapter 6 of [9]. The main argument from the quantum mechanical consideration is that at high $\boldsymbol{B}$ fields, back-scattering of the edge states is suppressed, because it becomes increasing difficult for states to scatter across the sample onto the opposite edge, which carries modes that travel in the reverse direction.

Landauer-Büttiker formalism

The Landauer-Büttiker formalism is a generalised method for calculating conduction in small quantum systems. In essence, it models electron transport as a problem of transmission probabilities. The simplest model is the two-terminal electron transport problem in a mesoscopic sample, where the sample is treated as a mesoscopic region that is linked by two perfect waveguides called leads, to two macroscopic conductors known as reservoirs. It is assumed that each reservoir provides thermally-equilibrated electrons which follow their own Fermi-

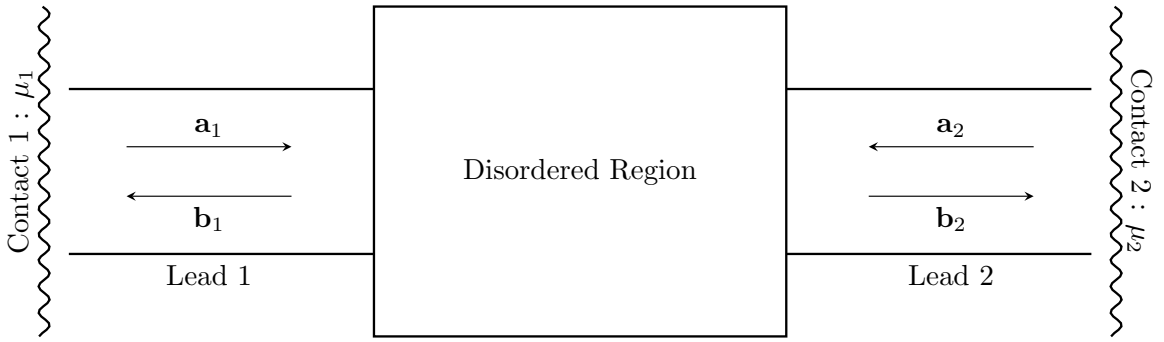


FIGURE 5.8 Two-terminal model of a mesoscopic sample. $\boldsymbol{a}_i = (a_{1i}, a_{2i}, \cdots, a_{ni})$ corresponds to the amplitude of the states entering the disordered region at lead i ($i = 1, 2$), with n being the label of each channel. Similarly, $\boldsymbol{b}_i$ is the vector of the amplitude of the reflected states.

Dirac distribution, uniquely described by chemical potential μ_i; and electrons propagate in the leads without scattering. Furthermore, electrons that impinge on the reservoirs are absorbed without reflection. As illustrated in Figure 5.8, electrons are injected into the sample from reservoirs on both sides. In the leads, current is carried by the eigenstates of the electron wavefunction, known as channels. The probability amplitudes of electron states incoming to the mesoscopic region in channels of lead i are represented by vector $\boldsymbol{a}_i$, and the amplitudes of states leaving the region are represented by vector $\boldsymbol{b}_i$. They can be reflected back to another channel in the same lead or transmitted to a channel in the opposite lead in scattering events that take place in the mesoscopic region. This process is summarised in one expression using a scattering matrix, $\boldsymbol{S}$ [9, 67]:

$$\boldsymbol{b} = \boldsymbol{S} \cdot \boldsymbol{a} \quad \Longrightarrow \quad \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{r} & \boldsymbol{t}' \\ \boldsymbol{t} & \boldsymbol{r}' \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{pmatrix}, \tag{5.10}$$

where $\boldsymbol{r}, \boldsymbol{t}$ are matrices of the reflection and transmission probability amplitudes of lead 1, and their primed variables are the counterparts of lead 2. Current conservation requires $\boldsymbol{b}^\dagger \boldsymbol{b} = \boldsymbol{a}^\dagger \boldsymbol{a}$, from which one may deduce $\boldsymbol{S}^\dagger \boldsymbol{S} = 1$ (i. e. $\boldsymbol{S}$ is unitary). It follows that

$$\boldsymbol{r}^\dagger \boldsymbol{r} + \boldsymbol{t}^\dagger \boldsymbol{t} = \boldsymbol{r}\boldsymbol{r}^\dagger + \boldsymbol{t}' \boldsymbol{t}'^\dagger. \tag{5.11}$$

The current in channel n travelling from contact 1 to 2 is

$$\begin{aligned} I_{1n} = \frac{2e}{L} \sum_k v(E) f(E - \mu_{1b}) T_n^{12} = \frac{2e}{L} \sum_k \frac{1}{\hbar} \frac{\partial E}{\partial k} f(E - \mu_1) T_n^{12} \\ = \frac{2e}{h} \int_0^\infty f(E - \mu_1) T_n^{12} \, \mathrm{d}E \,, \end{aligned} \tag{5.12}$$

where L is the length of the sample, $v(E)$ the drift velocity of electrons, $f(E)$ the Fermi-Dirac distribution, and T_n^{12} the probability of transmission from lead 1 to 2 in channel n. In the last equality of Equation 5.12, the discreet sum over all states is replaced by the continuous integral: $\sum_k \to \frac{L}{2\pi}\int \mathrm{d}k$, in the case of very large density of states. A similar expression exists for the current that travels from contact 2 to 1. The total current across the sample is obtained by summing over all the current carrying channels from both contacts:

$$I = \frac{e}{h}\sum_n \int_0^\infty \left[f(E-\mu_1)T_n^{21} - f(E-\mu_2)T_n^{12}\right]\mathrm{d}E\,. \tag{5.13}$$

By definition, the transmission probabilities are given by the mod-square of amplitudes: $T_n^{12} = (\boldsymbol{t}^\dagger \boldsymbol{t})_{nn}$ and $T_n^{21} = (\boldsymbol{t'}^\dagger \boldsymbol{t'})_{nn}$. Furthermore, by applying the trace operator to Equation 5.11 and noting that $\mathrm{Tr}(\boldsymbol{r}^\dagger \boldsymbol{r}) = \mathrm{Tr}(\boldsymbol{r}\boldsymbol{r}^\dagger) = \sum_n |r|^2_{nn}$, we have $\sum_n T_n^{21} = \sum_n T_n^{12} = \mathrm{Tr}(\boldsymbol{t}^\dagger \boldsymbol{t})$. Substituting the trace and the equation above becomes

$$I = \frac{e}{h}\int_0^\infty \mathrm{Tr}(\boldsymbol{t}^\dagger \boldsymbol{t})\left[f(E-\mu+eV_1) - f(E-\mu+eV_2)\right]\mathrm{d}E = \frac{e^2}{h}\mathrm{Tr}(\boldsymbol{t}^\dagger \boldsymbol{t})(V_1 - V_2), \tag{5.14}$$

where the equality is obtained using the result about the integral of the difference between two Fermi-Dirac distributions, which was discussed in Chapter 3 (see Figure 3.1). The result is known as the Landauer current formula. It can be rewritten to give the conductance across a two-terminal mesoscopic sample:

$$G = \frac{I}{(V_1 - V_2)} = \frac{e^2}{h}\cdot \mathrm{Tr}(\boldsymbol{t}^\dagger \boldsymbol{t}) \tag{5.15}$$

Integer quantum Hall effect

The sample resistance of integer quantum Hall effect can be calculated by combining the Landauer-Büttiker formalism with the edge channel model. Figure 5.9 shows a sample Hall bar in a typical four-terminal measurement setup. The conduction is carried by N edge states, and the total current passing from the source to drain contact is I, we can write the following group of equations using the result from Equation 5.14:

$$\begin{aligned} I &= \frac{e^2}{h}N(V_\mathrm{S} - V_4) & 0 &= \frac{e^2}{h}N(V_1 - V_\mathrm{S}) & 0 &= \frac{e^2}{h}N(V_3 - V_\mathrm{D}) \\ -I &= \frac{e^2}{h}N(V_\mathrm{D} - V_2) & 0 &= \frac{e^2}{h}N(V_2 - V_1) & 0 &= \frac{e^2}{h}N(V_4 - V_3) \end{aligned} \tag{5.16}$$

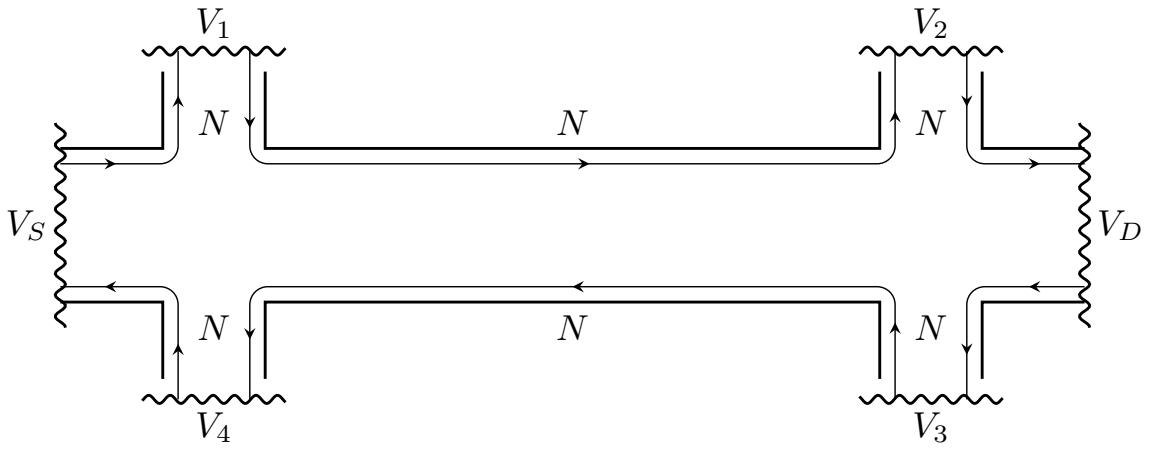


FIGURE 5.9 The model of edge current flow in a four-terminal quantum Hall experiment. A current I is passed between the source and drain contacts (sub-label S and D) by N edge channels. The probes at either side of the sample are used to measure longitudinal and transverse voltages of the sample.

solving the equations yields $V_1 = V_2$ and $V_3 = V_4$, which leads to $R_{xx} = 0$. Additionally,

$$I = \frac{e^2}{h} N(V_1 - V_4) \qquad I = \frac{e^2}{h} N(V_2 - V_3), \tag{5.17}$$

which leads to $R_{xy} = h/Ne^2$. These results are in good agreement with experimentally observed values of the longitudinal resistance plateaux; they also give the correct value of the transverse resistance, which is zero where the longitudinal resistance is on the plateaux. The number of edge channels, N, is the same as the number of populated Landau levels.

While the above discussion calculates the values of the resistance plateaux accurately, it does not explain why they appear in the first place. To answer this, we need to consider disorder broadening in the Landau levels. This refers to the lifting of degeneracy in the Landau levels by impurities disorder. As depicted in Figure 5.6(b), the effect of sample impurities is that the infinitely sharp Landau levels are broadened to cover a wide range of energy. The number of states in the broadened levels is the same as the degeneracy of the original levels, eB/h, and the centre-to-centre energy spacing between adjacent levels remains $\hbar\omega_c = \hbar eB/m$. Consequently, two things happen when we increase B: first, more states can fit into a single levels; second, the spacing between the levels increases. In a quantum Hall measurement the carrier density of the sample is fixed. Therefore, with increasing B, the Landau levels effectively rise above the Fermi level. The plateaux are observed when the Fermi level is between Landau levels: until a new level rises above E_F, the number of occupied levels, N, does not change, and therefore a constant resistance is observed.

Shubnikov–de Haas effect

The Shubnikov–de Haas (SdH) effect refers to the oscillation of longitudinal conductance between zero and some finite value in a Hall measurement. It was discussed in the last section that $R_{xx} = 0$ in a Hall bar when the R_{xy} is on a plateau. Since

$$\boldsymbol{\sigma} = \boldsymbol{\rho}^{-1} = \begin{pmatrix} 0 & \rho^{-1} \\ \rho^{-1} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}, \tag{5.18}$$

we see that $\sigma_{xx} = 0$ occurs when R_{xy} is on a plateau (therefore $\rho_{xx} = 0$). To see when σ_{xx} is finite, refer back to the discussion on Figure 5.7(a): states are likely to drift along closed paths when they are close to a peak or trough of confinement potential in the material bulk. This means in a broadened Landau level, states at the extreme ends of the energy spectrum are localised. There exists a narrow window of energies between the localised extremes, where it is possible for bulk states to be extended across the sample. This is represented by the shaded and empty regions in Figure 5.6(b). When the magnetic field is swept, the Fermi level moves relatively through the Landau levels. Whenever it is between the levels, or in the range of localised energies, $\sigma_{xx} = 0$. When it lands in the middle of a level, the extended states in the bulk result in finite conductance.

The oscillation period of the SdH effect can be calculated using the degeneracy of the Landau levels. Suppose a sample has total electron density n which is distributed into ν Landau levels in magnetic field B_i.

$$n = \frac{N}{A} = \nu \frac{eB_i}{h} \quad \Longrightarrow \quad \frac{1}{B_i} = \frac{e}{nh}\nu \quad \Longrightarrow \quad \frac{1}{B_{i+1}} - \frac{1}{B_i} = \Delta\left(\frac{1}{B}\right) = \frac{e}{nh} \tag{5.19}$$

The last expression shows that the conductance peaks occur periodically in inverse magnetic field, whose periodicity can be used to determine the carrier density n.

5.2.1 Wafer characterisation

The previous section showed that the Hall measurement can be used to measure the transport characteristics of a wafer such as the carrier density and mobility. This section describes the experimental setup we used to conduct the characterisation measurements on the GaAs/AlGaAs wafer for the electron tunnel device.

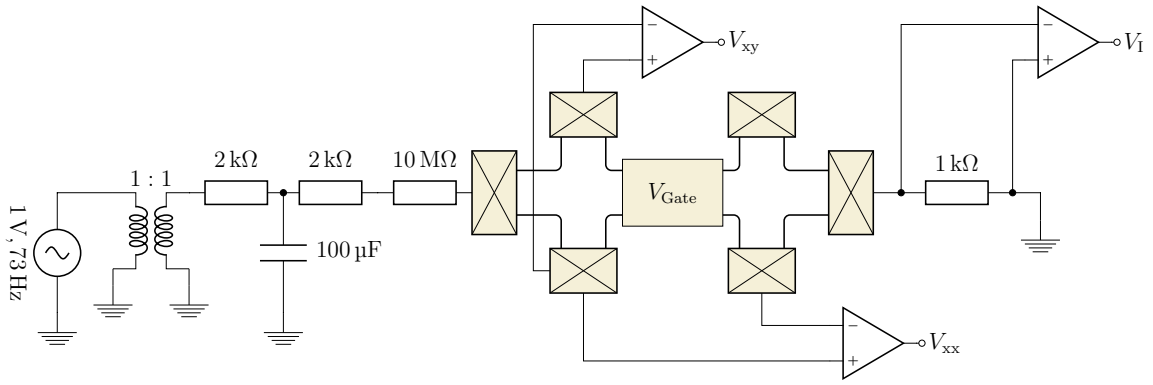


FIGURE 5.10 Schematic of the circuit setup of a four-terminal Hall measurement. The outputs, V_{xx}, V_{xy} were measured by two SR830 LIAs before being amplified by differential amplifiers built by the electronic workshop of the Cavendish laboratory. V_I was passed through a FEMTO current amplifier and measured by a model 7265 LIA.

Experimental setup of the four-terminal measurement

A Hall measurement involves a four-terminal setup, which measures the transverse and longitudinal resistances of a Hall bar. As shown in Figure 5.10, the Hall bar is a rectangular sample with the current source and drain contacts on opposite ends of the long axis of the sample. Four probe contacts are located along the long edges. The bulk of the Hall bar is covered by a continuous surface gate which can be used to control the carrier density of the wafer by varying the voltage V_{gate}. The measurement is carried out at low temperature in a magnetic field perpendicular to the Hall bar plane. In our work, this was achieved by mounting the sample inside a pumped ^{4}He cryostat (designated as the 10-Tesla). The surface gate was connected to a digital-to-analogue converter (DAC) which controls V_{gate}. An SR830 lock-in amplifier (LIA) was used to supply a 1 V sinusoidal voltage at 73 Hz across the sample. It was also used, along with two other LIAs, to measure the voltage response across the probe contacts, as well as the current through the sample. The AC output was passed through a 1 : 1 transformer to eliminate any DC offset, and a low-pass R-C filter (with 2 kΩ and 100 µF) to eliminate any sudden spikes in voltage. A 10 MΩ resistor was connected in series to the sample to limit the AC current. Since the sample resistance is significantly lower than the resistor, the excitation current can be considered as a constant 0.1 µA (zero to peak) AC signal. Using a differential amplifier (marked by V_I), the current was measured by the voltage drop across a small resistor (1 kΩ) which was connected in series to the sample. The transverse and longitudinal voltage drops across the hall bar was measured by two voltage amplifiers marked as V_{xy} and V_{xx}, respectively. The output of the three amplifiers were connected to separate LIAs that were synchronised to the input signal. An LIA is an instrument which measures the amplitude and phase of an oscillatory signal over a narrow bandwidth. By measuring V_{xy} and V_{xx}

at the excitation frequency, significant noise reduction can be achieved, as the LIA rejects the signal components outside the bandwidth window. The excitation frequency was chosen to be a prime number to minimise the chance of interference, particularly mains pick up, which occurs at 50 Hz and its integer multiples. Data gathering and instrument control are implemented by a dedicated language, CryoMeas, written by Prof. C. J. B. Ford in LabVIEW.

Results and Analysis

Figure 5.11 shows the result of a Hall measurement on a sample fabricated using wafer W938. The experiment was run at the base operating temperature of the 10-Tesla, which was 1.2 K. The result consists of the longitudinal and transverse resistance of the sample ($R_{xx} = V_{xx}/V_{\mathrm{I}}$ and $R_{xy} = V_{xy}/V_{\mathrm{I}}$, using the amplifier output shown in Figure 5.10) when the magnetic field was swept between 0 and 5 T. Each set of results contains two sweeps, one recorded when the field was ramped up, and the other during the ramp-down. The different sets differ by the value of the gate voltage, V_{gate}, which varied from 0.4 V to −0.6 V with steps of −0.05 V.

As explained in the earlier subsection on the Shubnikov–de Haas effect, the oscillation in the longitudinal conductance (equivalently, resistance) of a Hall measurement can be used to determine the electron density. Rearranging Equation 5.19, one obtains

$$n = \frac{2e}{h} \frac{1}{\Delta\left(\frac{1}{B}\right)}. \tag{5.20}$$

In practice, the term $[\Delta(1/B)]^{-1}$ was obtained by performing a fast Fourier transform (FFT) of R_{xx} in the inverse B-space. The numerical processing was done in Matlab: As the FFT algorithm requires evenly-spaced input, the B-inverted R_{xx} was interpolated on a grid of even number of points. A windowing function (specifically the Hanning function) was applied to the interpolated data to reduce spectral leakage (i. e. the broadening of FFT peaks in the frequency domain when the time domain input is not strictly an integer multiple of the signal frequency). The R_{xx} data was then passed to the FFT algorithm to compute the frequency components. Finally, the FFT output was converted into density by multiplying the conversion factor $2e/h$.

Figure 5.11b shows the electron density obtained from the FFT of the longitudinal resistance. Two fundamental frequencies of oscillations are revealed by the plot, each corresponding to a 2DEG layer in the sample. It can be seen that the 2DEG densities deplete linearly as the gate voltage is made more negative: From 0.4 V to −0.35 V, n_{upper} decreases continuously while n_{lower} is constant. This is due to screening: as long as the upper 2DEG remains populated, the electric field of the gate is shielded by the free electrons and cannot reach the

lower 2DEG. We can approximate the variation of the 2DEG densities in response to the gate voltage by two linear series in Figure 5.11c. A third curve, labelled by n_{Hall}, shows the combined electron density of the sample as calculated from the classical Hall effects: according to Equation 5.5, the density of the sample can be obtained from the gradient of the Hall resistivity as $n_{\text{Hall}} = \left(e \cdot \mathrm{d}\rho_{xy}/\mathrm{d}\boldsymbol{B}\right)^{-1}$. Figure 5.11c shows that this result agrees well with the density obtained from FFT in the region $V_{\text{gate}} < -0.4\,\text{V}$, where the upper 2DEG was fully depleted, and can be used to verify the validity of the FFT result.

The electron mobility of the sample can be calculated using the longitudinal conductivity. By definition, the mobility μ is related to the drift velocity $\boldsymbol{v}$ of the charge carriers and the electric field $\boldsymbol{E}$ by

$$\boldsymbol{v} = \mu \boldsymbol{E}. \tag{5.21}$$

Since the current density is given by $\boldsymbol{J} = ne\boldsymbol{v}$ and we have shown in earlier section that $\boldsymbol{J} = \sigma \boldsymbol{E}$, in the absence of magnetic field, we have

$$\sigma_{xx} = ne\mu, \tag{5.22}$$

where n is the carrier density and e the electric charge. Figure 5.11d shows two series of longitudinal conductivity values at $\boldsymbol{B} = 0$, calculated as $\sigma_{xx} = 1/R_{xx}/\text{number of squares}$. The number of squares is equal to the ratio between the length and width of the conduction region of the Hall bar. Its value is 8.25 for the test sample. The two series were calculated using the first and last values of $\boldsymbol{R}_{xx}$ from magnetic field up-sweeps and down-sweeps, respectively, and are in good agreement. The up-sweeps series was used to calculate the total mobility of the sample, using the expression:

$$\mu_{\text{eff}} = \frac{\sigma_{xx}}{e\left(n_{\text{upper}} + n_{\text{lower}}\right)}, \tag{5.23}$$

where the carrier densities were taken to be functions of the gate voltage as depicted in Figure 5.11c and the result is plotted as the blue series in Figure 5.11e.

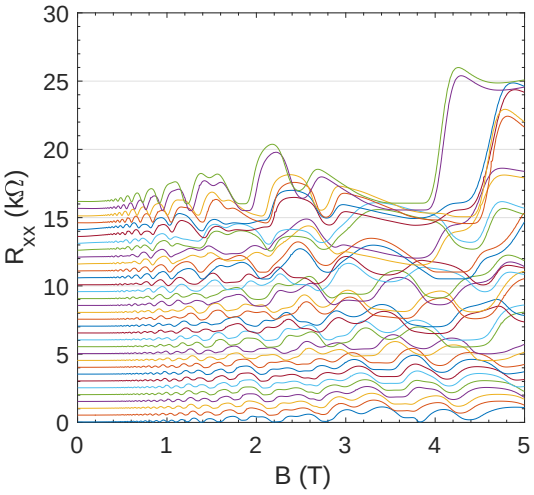
The mobilities of the individual layers can be worked out by referring to the two carrier model [49], which states that the effective total mobility of a two layer sample is given by:

$$\mu_{\text{eff}} = \frac{n_{\text{upper}}\mu_{\text{upper}} + n_{\text{lower}}\mu_{\text{lower}}}{n_{\text{upper}} + n_{\text{lower}}}. \tag{5.24}$$

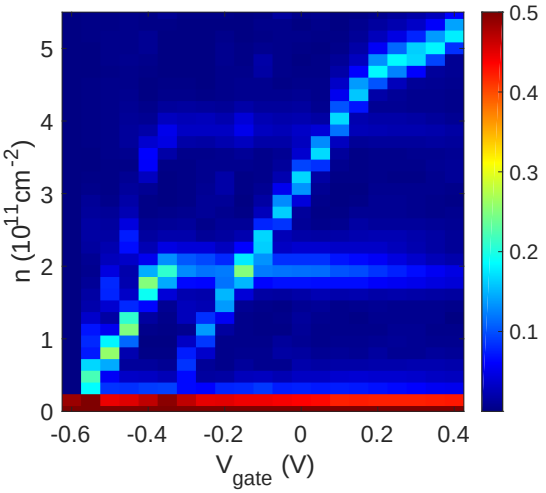
Figure 5.11c shows that the upper layer is fully depleted when V_{gate} is more negative than $-0.35\,\text{V}$. Since Equation 5.24 is simply $\mu_{\text{eff}} = \mu_{\text{lower}}$ when n_{upper} vanishes, the mobility of the lower layer is equal to the effective mobility of the sample when the upper layer is fully

depleted. The red series of points in Figure 5.11e was obtained by plotting Equation 5.23 with $n_{\text{upper}} = 0$. According to the argument above, the overlapping points of the red and blue series correspond to the mobility values of the lower layer. As the gate voltage is raised above the critical value of −0.35 V, the lower layer is fully populated and n_{lower} is constant. We assume μ_{lower} is constant as well in this voltage range, and its value equal to the maximum of μ_{lower}. With this assumption, we can calculate the mobility of the upper layer by rearranging Equation 5.24 to solve for μ_{upper}. Figure 5.11f shows the mobilities of the individual layers thus obtained as functions of the carrier density. It can be seen that electrons in the upper layer have mobilities 211 $\text{m}^2/(\text{V s})$ while in the lower layer the value is 210 $\text{m}^2/(\text{V s})$, given the layers are fully populated. Note also that the maximum carrier density of the upper layer is $5.04 \times 10^{11}\ \text{cm}^{-2}$ and the that of the lower layer is $1.97 \times 10^{11}\ \text{cm}^{-2}$.

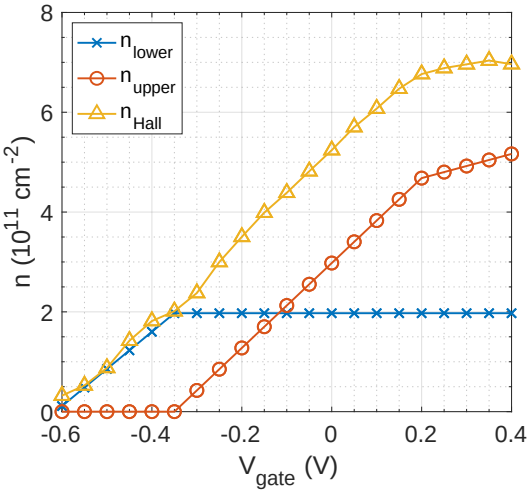
The same process as outlined above was applied to a test sample made from wafer W-939. Figure 5.12 shows the result and analysis of the measurement. The upper 2DEG has a maximum carrier density of $5.04 \times 10^{11}\ \text{cm}^{-2}$ and mobility of 136 $\text{m}^2/(\text{V s})$, and the lower 2DEG has a maximum density of $2.2 \times 10^{11}\ \text{cm}^2$ and a mobility of 162 $\text{m}^2/(\text{V s})$. For data on wafer C2617, see Chapter 3 of [44].



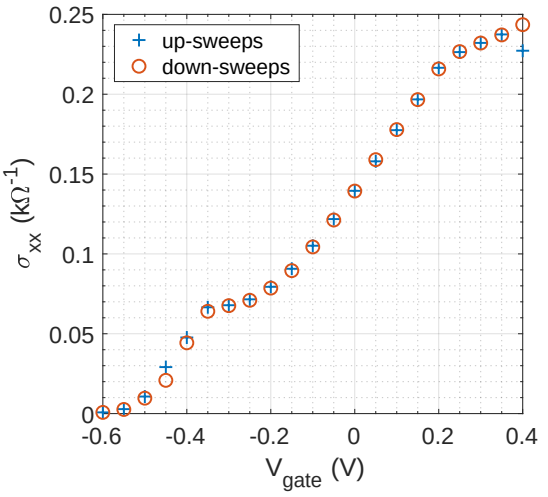
(a) Longitudinal resistance vs B



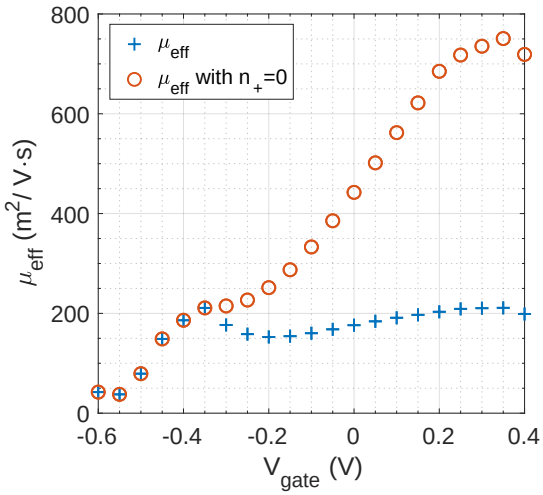
(b) Fourier transform of R_{xx} in B^{-1}



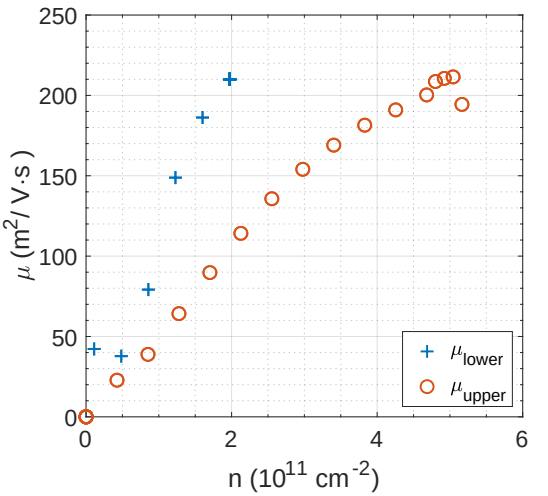
(c) Carrier densities vs gate voltage



(d) Longitudinal conductivity ($B = 0$) vs V_{gate}

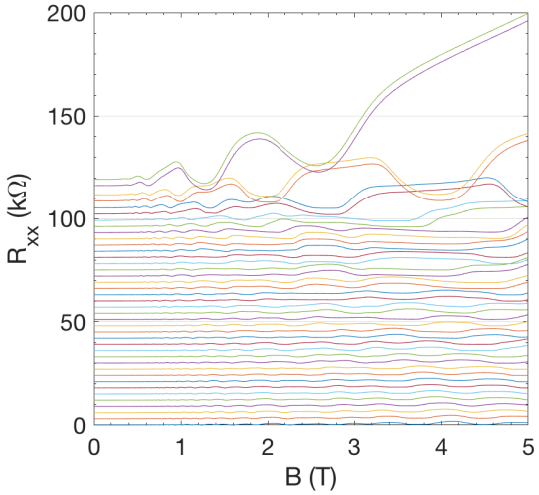


(e) Effective mobility vs V_{gate}

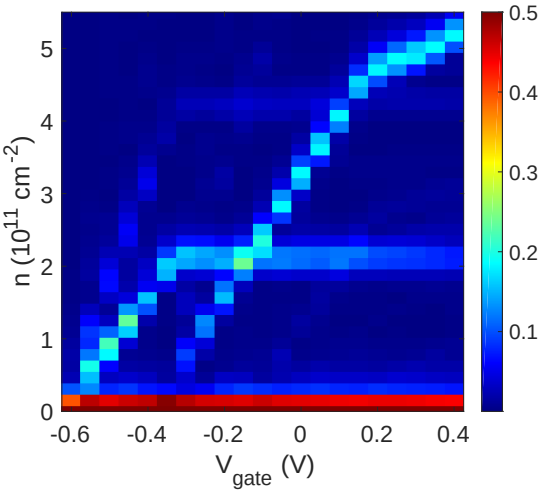


(f) Mobilities vs density of individual layers.

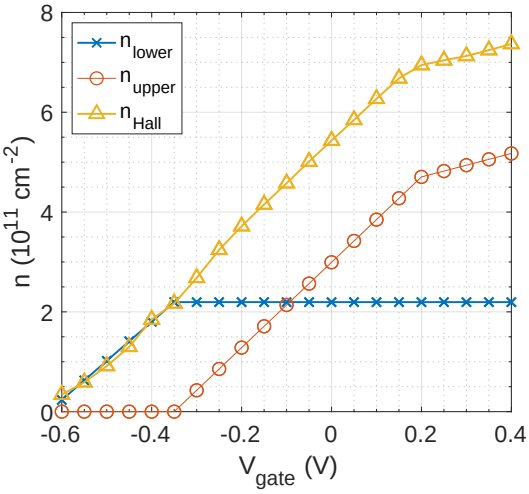
FIGURE 5.11 Results and analysis of Hall measurement of wafer W938, taken at 1.5 K. Adjacent sets of sweeps in (a) are offset by 0.5 kΩ for better presentation.



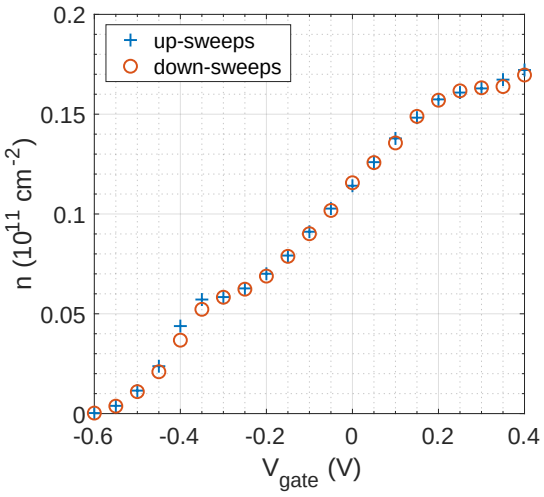
(a) Longitudinal resistance vs B



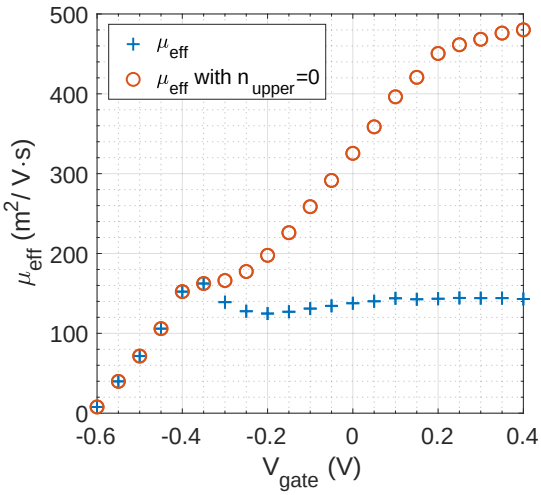
(b) Fourier transform of R_{xx} in B^{-1}



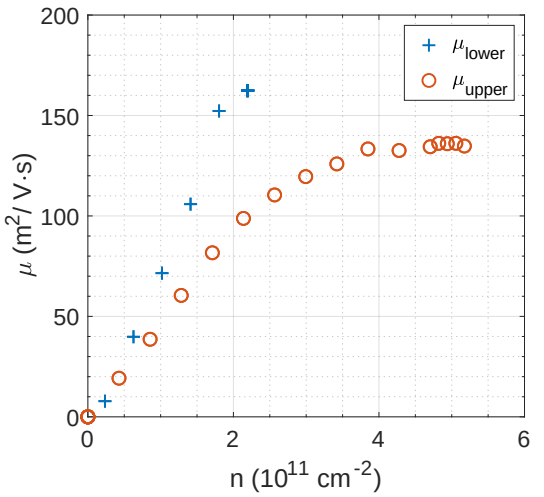
(c) Carrier densities vs gate voltage



(d) Longitudinal conductivity ($B = 0$) vs V_{gate}



(e) Effective mobility vs V_{gate}



(f) Mobilities vs densities of individual layers.

FIGURE 5.12 Results and analysis of Hall measurement of wafer W-939, taken at 1.5 K.

Chapter 6

Measurement of energy-momentum-resolved tunnelling

6.1 Introduction

As discussed in the previous chapter, the magneto-resonance tunnelling current is proportional to the convolution of the spectral functions of the double quantum wells (Equation 3.8), and the measurement of the tunnelling conductance can be used to resolve the spectral functions in energy-momentum space. In this chapter we discuss the experimental details of the project and make preliminary presentation of the gathered data without in depth analysis.

6.1.1 Layout of the tunnelling spectrometer device

Figure 6.1a shows a not-to-scale schematic that illustrates the structure of the surface gates. With the exception of the source/drain contacts, all surface gates are Schottky contacts. Their main function is to create an array of 1D channels by depleting the upper quantum, where electrons can tunnel into/out of the lower quantum well. Figure 6.1b is an illustration that shows the flow of electrons when the device is in the tunnelling configuration. We describe briefly below each surface gate and its function, before discussing their operation details in the following sections.

Source/drain contacts The Ohmic contacts labelled source and drain are connected to the double quantum wells. They are used to provide the DC bias (modulated with an AC modulation voltage for phase sensitive measurement) which controls the offset in the Fermi energies between the two quantum wells. As a convention, the positive pole of the DC bias was always connected to the Ohmic contact on the side of the barrier gate in our experiments.

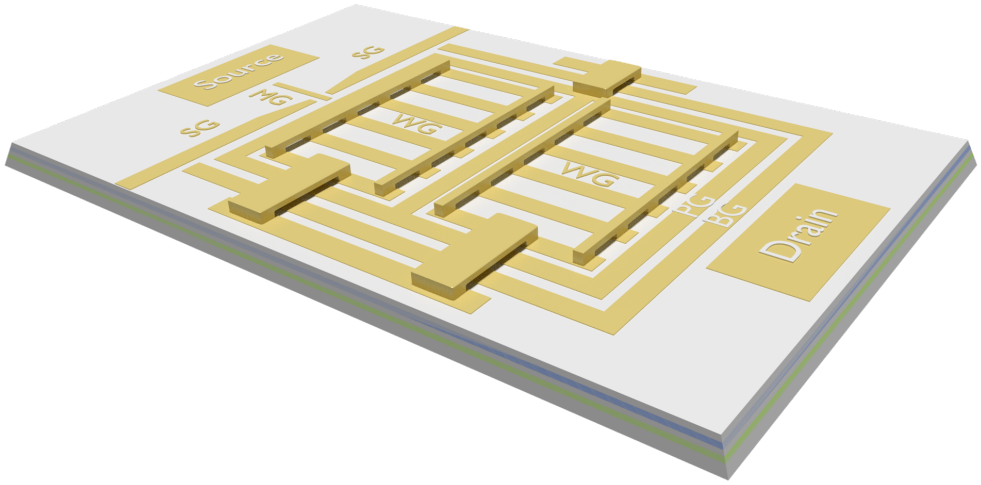
Split-gate/mid-gate and barrier gate In order to facilitate tunnelling between double quantum wells, one requires the source drain contacts to independently connect to each of the layer. In Chapter 2 we looked at several resonance tunnelling devices. All of them relied on the use of a depletion gate on the bottom surface of the sample to cut of the lower layer from one of the Ohmic contacts. In this work, we adopted the design proposed by Nield [66] that uses a split-gate/mid-gate arrangement on the top surface instead of the back gates, which greatly reduces the complexity of device fabrication. The split gate is a single gate that cuts across the mesa and has an aperture in the middle, whereas the mid gate is positioned in the middle of the aperture. These gates function in tandem to cut off the bottom quantum well from the near Ohmic contact while maintaining connection to the top well. On the opposite side of the mesa, connection to the top well is cut off with the use of a barrier gate, which is U shaped and surrounds the entire tunnelling area.

Wire gate array The wire gate array consists of a lattice of EBL fabricated repeating units (∼500 in total) which has open-ended channels in between. They are connected at both ends by air-bridges. Their function is to provide the confinement potential to modulate the upper well into a lattice of 1D channels.

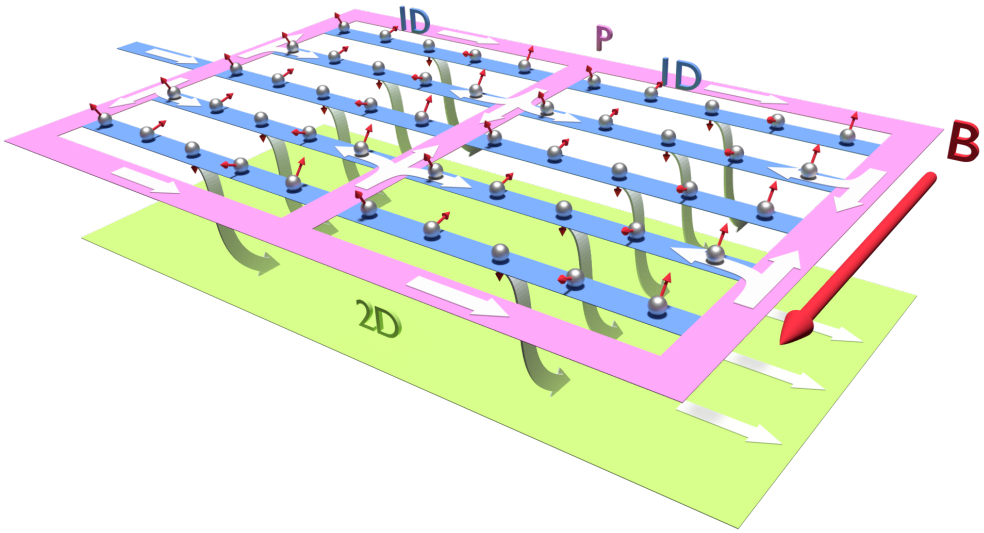
Parasitic gate There are regions in the tunnelling area which lie outside the 1D channels, but nonetheless under go tunnelling. In Figure 6.1b this area the pink coloured region that surrounds the 1D channels (blue region). We refer to this area as the parasitic region. A gate on top of this region provides control over its electron density. This is the parasitic gate.

6.1.2 Measurement circuit

The circuit for measuring the tunnelling conductance is similar to that used for wafer characterisation presented in Section 5.2.1, but differs in being a two-terminal instead of a four-terminal setup. As shown in Figure 6.2, the current across the sample is generated by the 1 V, 73 Hz sinusoidal output voltage of the measuring lock-in amplifier (SIGNAL RECOVERY model 7265). The transformer and the low-pass filter in series with the AC source are identical to those of the four-terminal setup, which are used for DC-decoupling and noise filtering. $1/10^4$ of the AC signal is passed to the sample via a voltage divider formed between resistors R_2 and R_1. The current through the sample is converted into a voltage signal by a Femto current pre-amplifier on 10^6 V/A gain, before being measured by the lock-in amplifier. Lock-in amplifiers measure voltage within a narrow band around a given frequency. In our setup the voltage detected by the lock-in is proportional to the change of the sample current δI in response to the AC excitation, which is a small ripple superimposed on the DC-bias across the



(a) Schematic of the tunnel device. The blue and green layers within the substrate represent the upper and lower 2DEG, respectively. The model is not to scale: on actual devices, the wire-gate array are tightly-packed ultra-fine e-beam features which cannot be distinguished by eye under optical microscope. Additionally, depending on the length of the repeating units, the WG array may contain more or less number of columns than that shown in the illustration. However the total tunnelling area is the same between device variations.



(b) Illustration of the current flow (indicated by the broad white arrows) during tunnelling operation. The upper 2DEG is heavily depleted into narrow channels by the surface gates, while the lower 2DEG remains unperturbed (2D) due to screening. The spheres represent electrons (with spin, represented by the red arrows) aligned into 1D lines. Theory predicts spin and charge density waves propagate at different velocities in 1D.

FIGURE 6.1 Schematics of the experimental device and the current flow within it during operation.

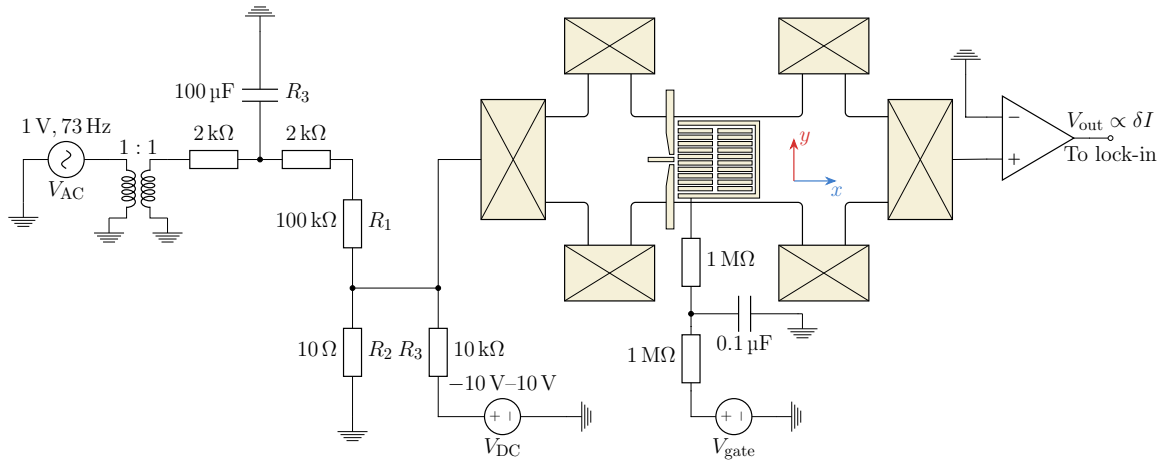


FIGURE 6.2 Circuit setup of the tunnelling conductance measurement. V_{AC} is the excitation voltage which drives the current across the sample. The 1:1 transformer is used to decouple DC bias from the AC source. The two 2 kΩ resistors and the 100 µF capacitor constitute a low-pass filter. V_{AC} is reduced by a fraction of R_2/R_3 before reaching the sample. V_{DC} is the DC-bias voltage which offsets the Fermi energies of the double 2DEGs in the sample. It is reduced by a fraction of R_2/R_3. The current is measured by a lock-in amplifier, after the signal is converted into a voltage by the current pre-amp at 10^6 V/A gain. V_{gate} is the voltage which controls the bias of surface gates. The two 1 MΩ resistors and the 0.1 µF capacitor form a gate filter which protects the gate from voltage surges such as ESD. In reality, each surface gate has its own dedicated voltage source and filter.

quantum wells. In measuring the change in current we are effectively determining the differential conductance of the sample, $\delta G = \delta I/\delta V$, where $V = V_{DC} + V_{AC}$ and $\delta V = V_{AC}$. The DC-bias is used to control the Fermi energy offset between the quantum wells. It is supplied via a $1/10^3$ voltage divider formed between resistors R_2 and R_3. The DC voltage source is a Yokogawa source measure unit (SMU) which ahas an output range of ±10 V. The bias voltages for the surface gates are supplied by 16-bit quad-channel digital to analogue converters (DACs). Each gate voltage connection is protected by filters made from two 1 MΩ resistors and a 0.1 µF capacitor, to minimise damage to the fine-feature surface gates from sudden current surges such as ESD. All measurements were conducted at cryogenic temperature: samples made from wafer C2617 were measured in a ^{3}He/^{4}He dilution refrigerator, while samples made from the W-series wafers were measured in a sorption-pumped ^{3}He cryostat.

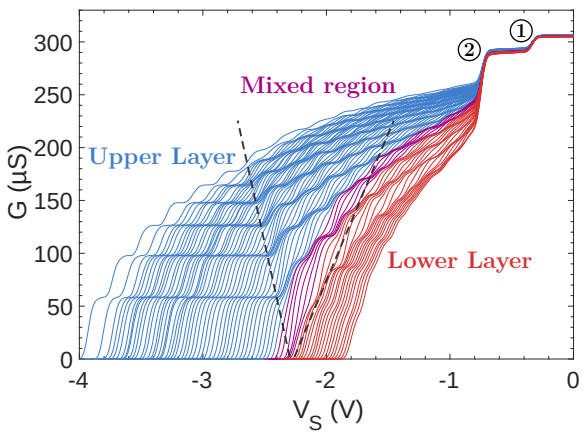
6.2 Independent Ohmic contacts

In order to induce a tunnelling current between the 2DEGs, it is necessary to restrict the conduction path such that the only way for a current to pass through the sample is via inter-layer tunnelling. As the impedance of tunnelling is much larger than the impedance of direct con-

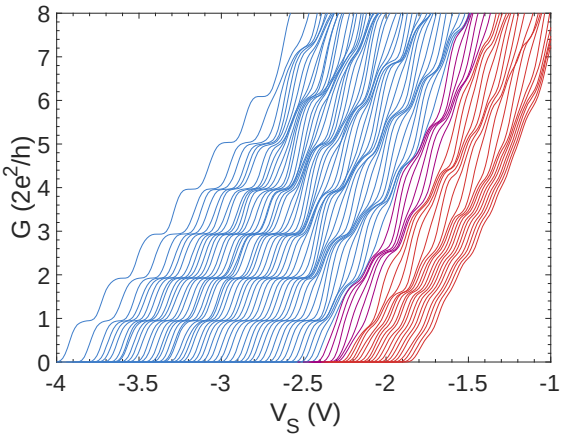
tacts (compare sweep ① and ⑥ in Figure 6.5), should the layers be electrically joined at any point, the current will bypass the tunnelling path via the connection. When Ohmic contacts are fabricated they are connected to both 2DEGs (Section 4.2.2). It is therefore necessary to introduce a mechanism to selectively cut off the connection between the Ohmic contacts and the 2DEGs, so that each Ohmic contact is connected to a different quantum well. This device configuration is called independent Ohmic contacts, which can be achieved by applying specific voltages on the SG/MG and BG.

The SG pair is used to cut off both 2DEGs while the MG induces a small conducting channel in the upper layer. The SG consists of a pair of 2 μm-wide gates separated by a 0.3 μm gap; the MG is a 0.1 μm-wide gate positioned in the centre of the SG gap. For SG/MG to work, first a negative bias needs to be swept on the SG until both 2DEG layers are cut off. Then a positive bias is swept on the mid-gate until a conducting channel is induced in the upper 2DEG. On the opposite end of the sample, the BG is used to cut off only the upper 2DEG, while leaving the lower layer unaffected. It is a 2 μm-wide gate, which surrounds the tunnel area of the device.

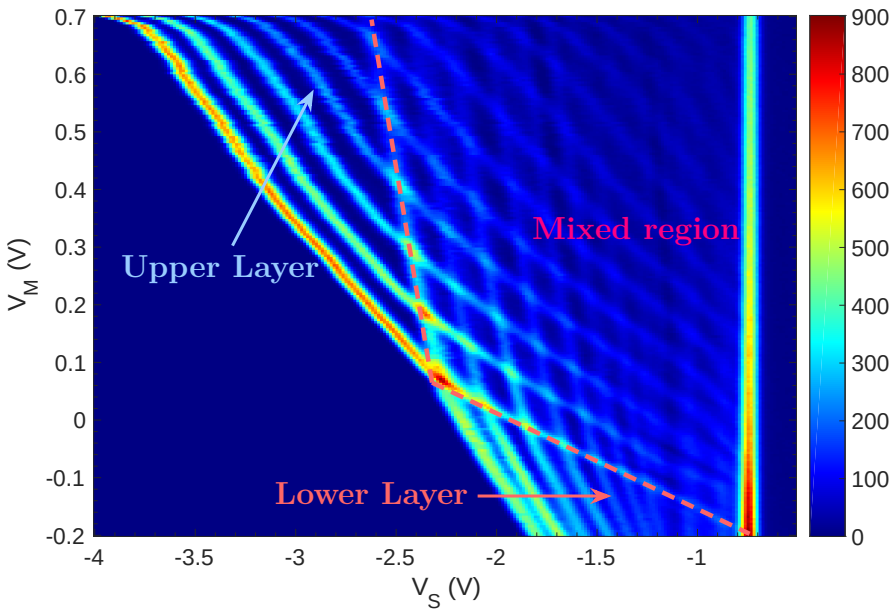
A calibration needs to be performed on each device to determine the range of SG/MG voltages for upper-2DEG-only connection. This was done by sweeping the SG with a negative voltage V_S while holding the mid-gate at a fixed bias V_M, and measuring the sample conductance. Figure 6.3 shows the results of such a measurement on a C2617 wafer test device conducted at 50 mK on a dilution fridge. V_M was varied between −0.2 V and 0.7 V, with an increment of 0.04 V between consecutive sweeps. Subfigure (a) shows the sample conductance as V_S is swept from 0 V until the device is pinched-off: initially, two conductance steps are observed at −0.27 V and −0.66 V, respectively (labelled ① and ②). This occurs as the electrons underneath the split-gates are depleted in the upper and lower layers. The conductance does not drop immediately to zero at ② because the SG gap is not pinched off. As the split-gate voltage is swept more negatively, the conductance is lowered further as the confinement potential in the gap becomes stronger. The conductance approaches 0 in a series of increasingly pronounced plateaux. This can be understood in terms of the quantised conductance in the Landauer-Büttiker formalism: according to Equation 5.14, each conducting channel (of unit transmission probability) contributes a conductance quantum of $G_0 = 2e^2/h = 77.5\,\mu\text{S}$ (the extra factor of 2 is due to spin degeneracy). The steps manifest because as the SG voltage becomes more negative, the number of conductance channels in the SG aperture is reduced one by one. Subfigure (b) shows the conductance can be normalised to integer multiples of G_0 after a series conductance of 285 μS is subtracted: suppose the measured conductance is G_m, the conductance associated with the SG aperture is G, and the series conductance is G_s. The formula of series conductance, $1/G_m = 1/G + 1/G_s$, implies that $G = G_m G_s/(G_s - G_m)$.



(a) Conductance vs. split-gate voltage

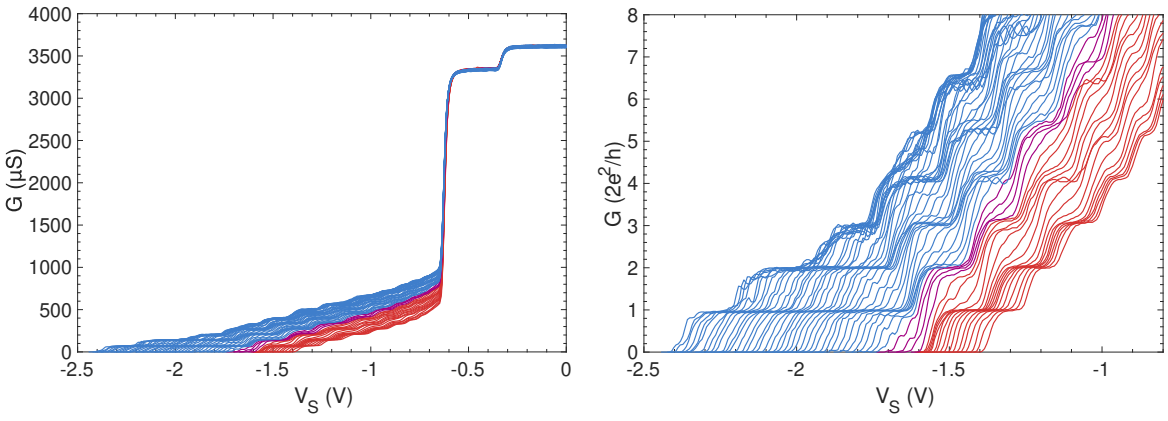


(b) Conductance in quantised unit



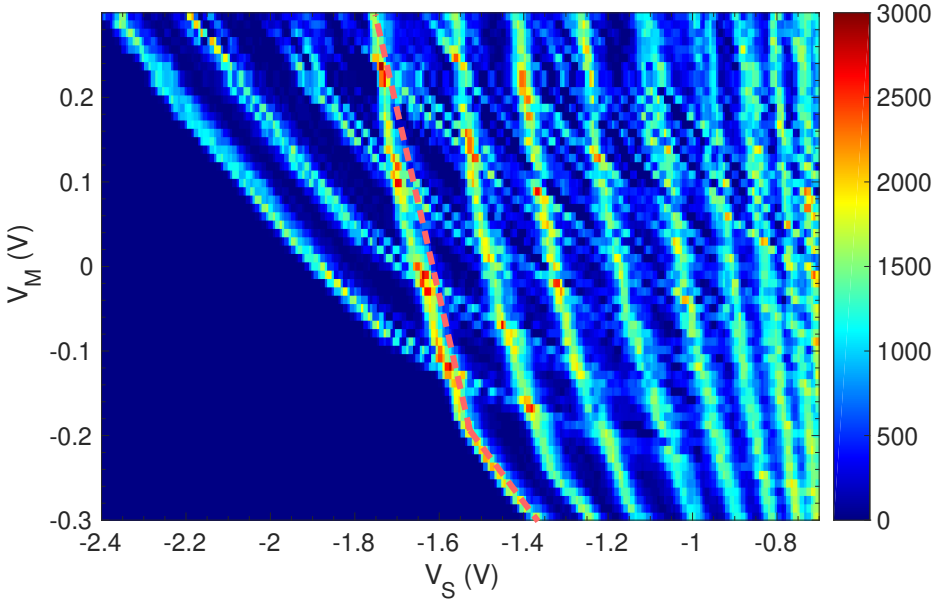
(c) Density map of conductance differential $\mathrm{d}G/\mathrm{d}V_s$

FIGURE 6.3 (a): Test device (wafer C2617) conductance, measured at 50 mK, as the split-gates are swept from 0 V until the conductance is cut-off. In each sweep the mid-gate voltage is reduced by 0.04 V from an initial value of 0.7 V to a final value of −0.2 V. To simplify presentation, only every third recorded sweep is displayed. The conductance drop at ① is resulted from the depletion of electrons in the upper QW directly beneath the spilt-gates; The drop at ② is caused by the depletion of both the upper and lower wells beneath the spilt-gates. (b): Sample conductance normalised to the conductance quantum $2e^2/h$ after a series conductance of 285 μS is subtracted from the result of (a). (c): Density map of the conductance differential $\mathrm{d}G/\mathrm{d}V_s$ plotted against the split-gate and mid-gate voltages, showing regions where the device conducts only in the upper, lower 2DEG, as well as in both layers.



(a) Conductance vs. split-gate voltage

(b) Conductance in quantised unit



(c) Derivative of conductance $\mathrm{d}G/\mathrm{d}V_s$

FIGURE 6.4 Test device (wafer W938) conductance, measured at 300 mK at different split gate voltage. A series conductance of 1900 μS was subtracted to produce subfigure (b).

The value G_s is determined iteratively by varying G_s until the plateaux of G line up with integer multiples of G_0. The series conductance can be attributed to the impedance of the Ohmic contact, as well as that of circuit components such as the transformer, filters and current preamp that are in series with the device. It can be observed from (a) and (b) that at certain voltages, the conductance plateaux have steps of size $2G_0$. This occurs when both 2DEGs contribute towards the total conductance. The region is demarcated by dashed lines, and is referred to as the mixed region. The sweeps that lie inside this area are coloured magenta. Outside this region, the conductance originates from channels in one particular 2DEG layer. It is easier to identify these regions by plotting the conductance differential to the swept variable ($\mathrm{d}G/\mathrm{d}V_s$) as a density map in the span of SG/MG voltages, as shown in (c). We select a mid-gate voltage that is well within the upper-layer-only to proceed with the tunnelling measurement. Figure 6.4 shows the equivalent result obtained from a test device fabricated on wafer W938. It was measured at 300 mK in a sorption-pumped ^{3}He cryostat.

6.2.1 Tuning for the tunnelling configuration

For the C2617 device that produced Figure 6.3, one can identify 0.5 V as a suitable mid-gate bias for upper-layer-only conduction. The device can be placed in the tunnelling configuration by a sequence of gate sweeps shown in Figure 6.5: In sweeps 1-4, the SG/MG are biased for the independent ohmic contact. First, the SG is swept negatively until the sample is pinched off. Next, while maintaining the SG bias, the MG is swept positively to the target value of 0.5 V. Then the SG voltage is lowered again to check that the sample can be pinched off completely. Afterwards, the SG voltage is raised to admit two conductance plateaux. This choice was due to the empirical observation that admitting three or more conductance quantum would result in noisier tunnelling signals later on. In sweep 5, the barrier-gate was swept until the sample is just pinched-off. This completes the independent Ohmic contact configuration. We then check that a tunnelling current can be induced in the device in sweep 6 by sweeping the WG array between positive and negative voltages. The change of density effected in the tunnelling region by this sweep should produce a conductance peak, which is identified as the tunnelling current.

6.3 Zero-DC-bias magnetic-field sweeps

Once the tuning sweeps detect the tunnelling conductance, more detailed studies of the device are carried out. First we measure the conductance at zero DC-bias by sweeping the magnetic field as the wire-gates are gradually pinched off. As explained in Figure 3.6, along

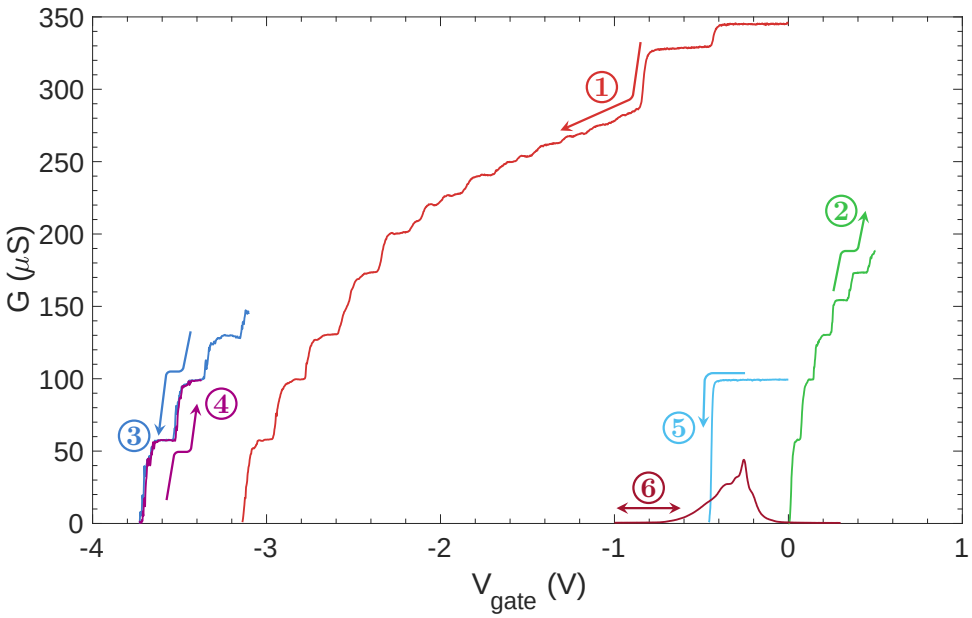
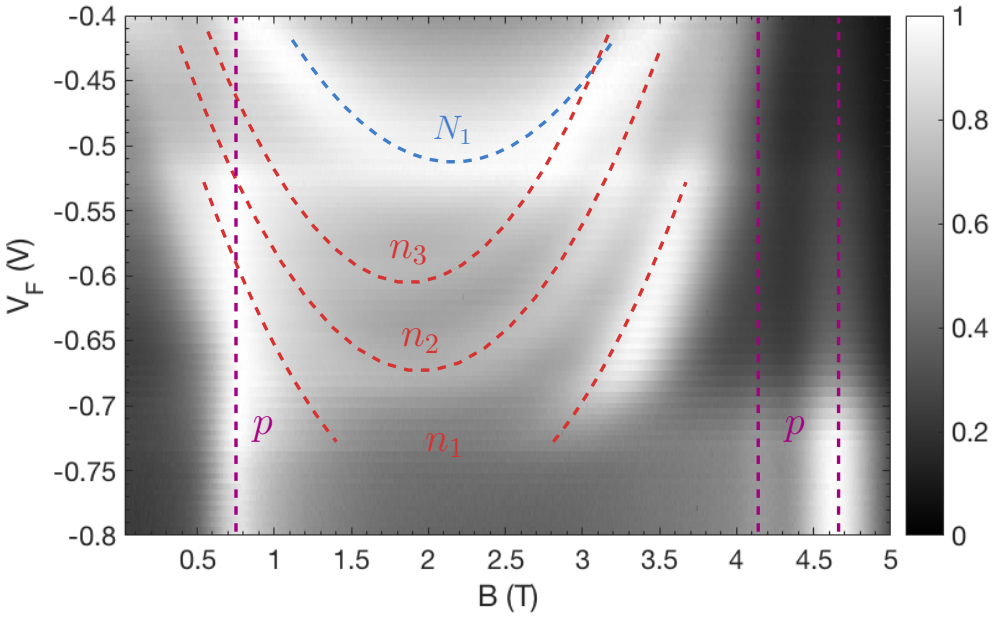


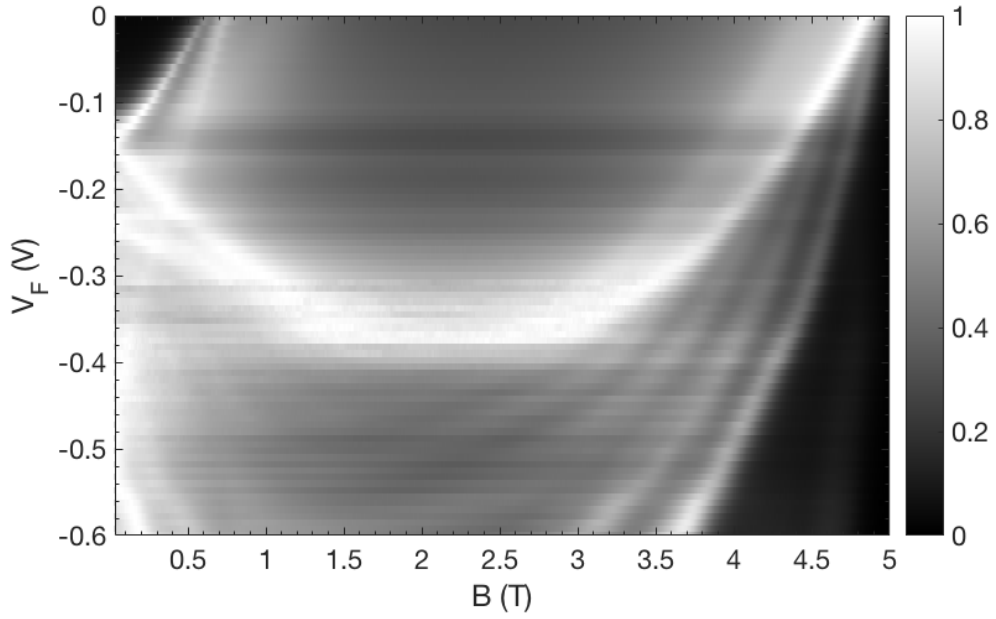
FIGURE 6.5 The sequence of voltage sweeps on different gates for tuning the C-2617 device to the tunnelling configuration, measured at 50 mK. The sample conductance is monitored as the various gates biases are swept in the direction of the arrow. ①: The split-gate bias is reduced until the sample is pinched-off. ②: The mid-gate is swept to a pre-determined (in reference to Figure 6.3c) voltage which allows only upper-layer conductance. ③: The split-gates are swept further until the device is pinched-off again. ④: The split-gate is relaxed to allow one or two conductance plateau(x) across the sample. ⑤: The barrier-gate is swept negatively until the sample conductance is just cut off. ⑥: The wire-gate array is swept between positive and negative voltages to verify that a non-zero conductance can be induced by varying the carrier density of the upper 2DEG. This conductance is attributed to the inter-layer tunnelling current.

every magnetic field sweep the maximum conductance occurs at two fields $B_{\mp}$, where the overlap of the spectral functions is the greatest. When V_{F} is made more negative, the electrons become increasingly depleted in the upper layer, reducing the effective radius of the 1D spectral function. According to Equation 3.24, this results in an ever smaller separation between $B_{\mp}$, which is shown in the result of Figure 6.6.

In the density plots, one can trace the conductance maximum along a series of curves, which demonstrates that the device is able to resolve tunnelling signals with multiple 1D/2D subbands. This shows that the tunnelling signals from the repeating units of the wire-gate array are coherent, indicating that the array does not suffer from significant lithographic irregularity, and the air-bridge array which connects them functions properly; it also reveals the number of 1D subbands one may expect the device to resolve at a particular V_{F}, as well as the most negative value V_{F} that can be applied before the wire-gate array pinches off. In Figure 6.6a, it can be observed that the 2D subband persists down to $V_{\mathrm{F}} \approx -0.52\,\mathrm{V}$. This



(a) Zero DC conductance of sample C2617



(b) Zero DC conductance of sample W938

FIGURE 6.6 Intensity maps of normalised tunnelling conductance plotted against the wire-gate voltage V_F and the magnetic field B. The conductance is normalised to the maximum of each sweep (along the B axis) to better show the conductance across the whole measurement range. In subfigure (a), dashed curves are plotted to highlight the various features on the density map. Blue: 2D subband; red: 1D subbands; magenta: parasitic conductance features.

is when electrons in the upper layer are not depleted strongly enough to form 1D channels. Below this value, three 1D subbands can be identified, with the last one cutting off at $V_F \approx -0.72$ V. Additionally, field-independent features (magenta) can be seen at $\boldsymbol{B} = 0.8$, 4.1 and 4.6 T, which are caused by parasitic tunnelling that was discussed in Section 6.1.1.

6.4 Full map of tunnelling conductance

Once a device is shown to be able to resolve 1D/2D subbands in the zero-DC-bias magnetic-field sweeps, we proceed with the full measurement of its tunnelling conductance in energy-momentum space. The measurement is carried out by holding the wire-gate voltage constant while sweeping the DC-bias voltage. After each V_{DC} sweep, the magnetic-field is incremented by a small amount, calculated automatically by CryoMeas based on the specified initial/final value and total number of sweeps. The DC sweep and the $\boldsymbol{B}$ increment are repeated until the full magnetic field range is covered. Figure 6.7 shows a collection of full V_{DC}–$\boldsymbol{B}$ maps measured at six increasingly negative wire-gate voltages (V_F), from a sample fabricated on the C2617 wafer (Sample I of the next section). Each map of the figure consists of 100 DC-bias sweeps between a voltage range of –10 to 10 mV, conducted at a rate of 1200 mV h^{-1}, in equal increments between magnetic field 0 to 6 T. Figure 6.8 shows the equivalent results from a sample fabricated on the W938 wafer (Sample II of the next section). Note that on preliminary examination, the tunnelling-conductance peaks resemble the parabolic dispersions plotted in Figure 3.5a. Significantly, negative differential conductance (see ④ in Figure 3.5b) can be observed in the top left corner of most maps between $V_{DC} = 2$ to 10 mV and $\boldsymbol{B} =$ 0.6 to 1.5 T. Each map contains two sets of parabolae: in Figure 6.7, the first set has zero-DC crossing points at ∼0.7 T and ∼3.5 T while the second set has crossing points at ∼0.65 T and ∼4.6 T; in Figure 6.8, the first set has crossing points ∼0.2 T and ∼3.7 T, and the second set has crossing points ∼0.8 T and ∼4.7 T. The two sets of parabolae are distinguished by the observation that the first one changes noticeably across the figures while the second one remains constant. Initially at the least-negative gate voltage, two subbands can be observed in the first set of parabolae of Figure 6.7, and three subbands can be seen in that of Figure 6.8. As the wire-gate voltage (V_F) is made more negative, the number of subbands is gradually reduced, and the crossing points of the subbands move closer together. Furthermore, the overall conductance of the maps diminishes, which is reflected in the decreasing limits of the colour bars. The dependency on V_F suggests that the first set of parabolae is associated with the tunnelling in the wire-gate array, and the trend is consistent with the expected behaviour of decreasing electron density of the upper 2DEG as the wire gates are increasingly depleted. The second set of parabolae, being independent of V_F, is produced by tunnelling outside the

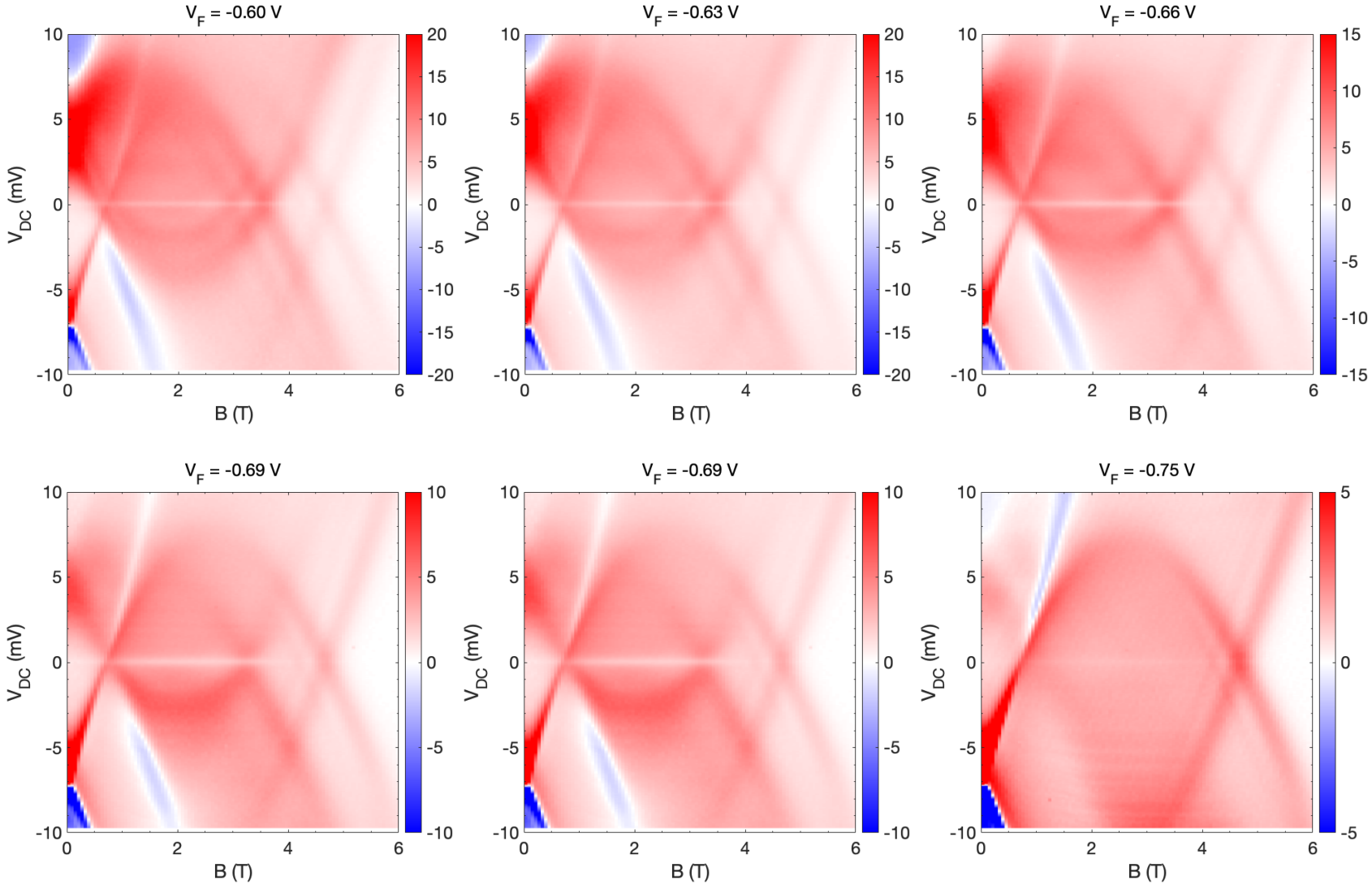


FIGURE 6.7 Density map of the tunnelling conductance (μS) in B–V_{DC} space, measured at different wire gate voltages (V_F) on Sample I, measured at 50 mK.

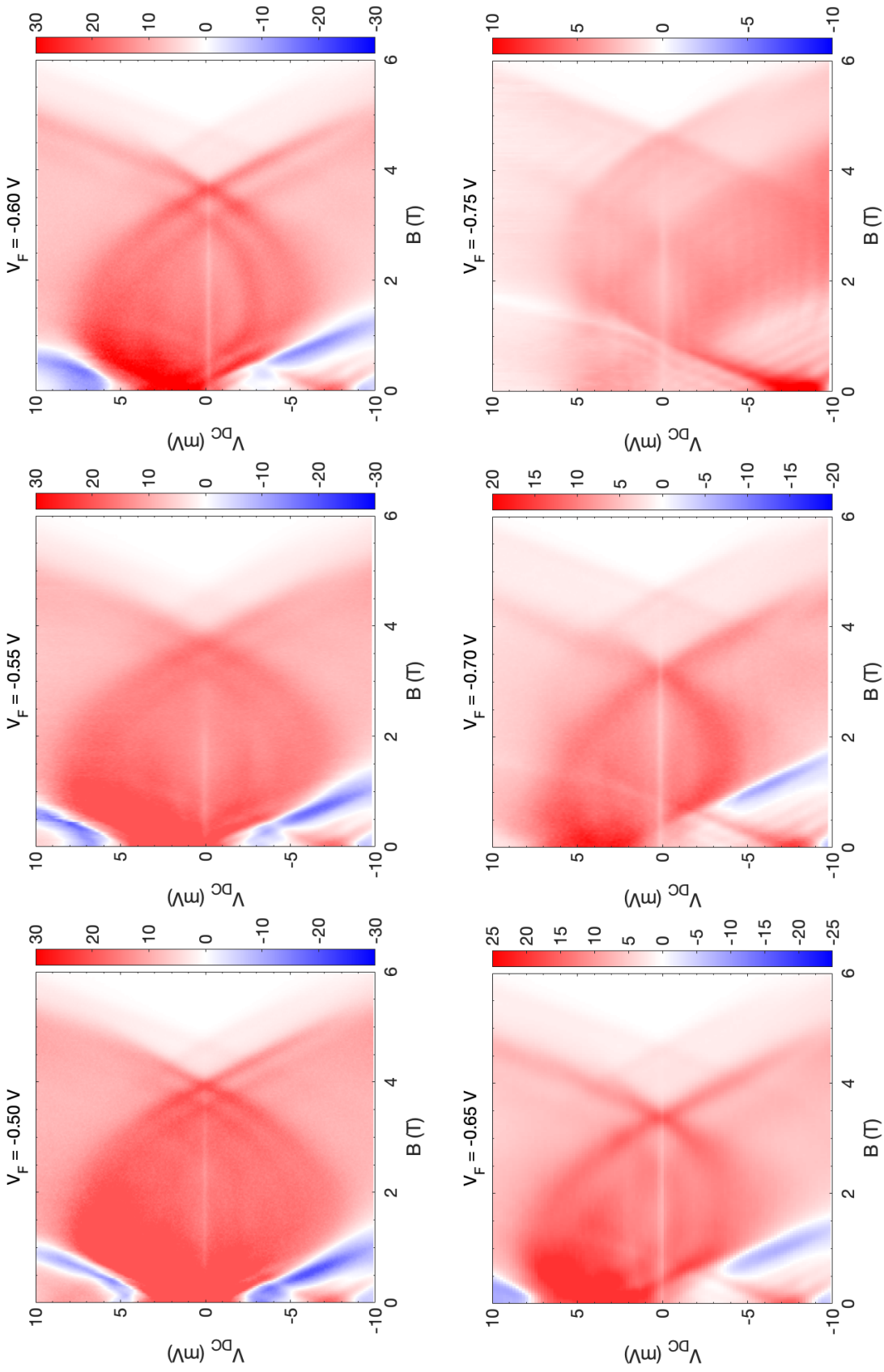


FIGURE 6.8 Density maps of the tunnelling conductance (μS) measured at different wire gate voltages (V_F) on a Sample II, measured at 300 mK.

wire-gate array. This is due to conductance from the 'parasitic' region of the sample, which was explained in the previous section on the zero DC-bias B sweeps.

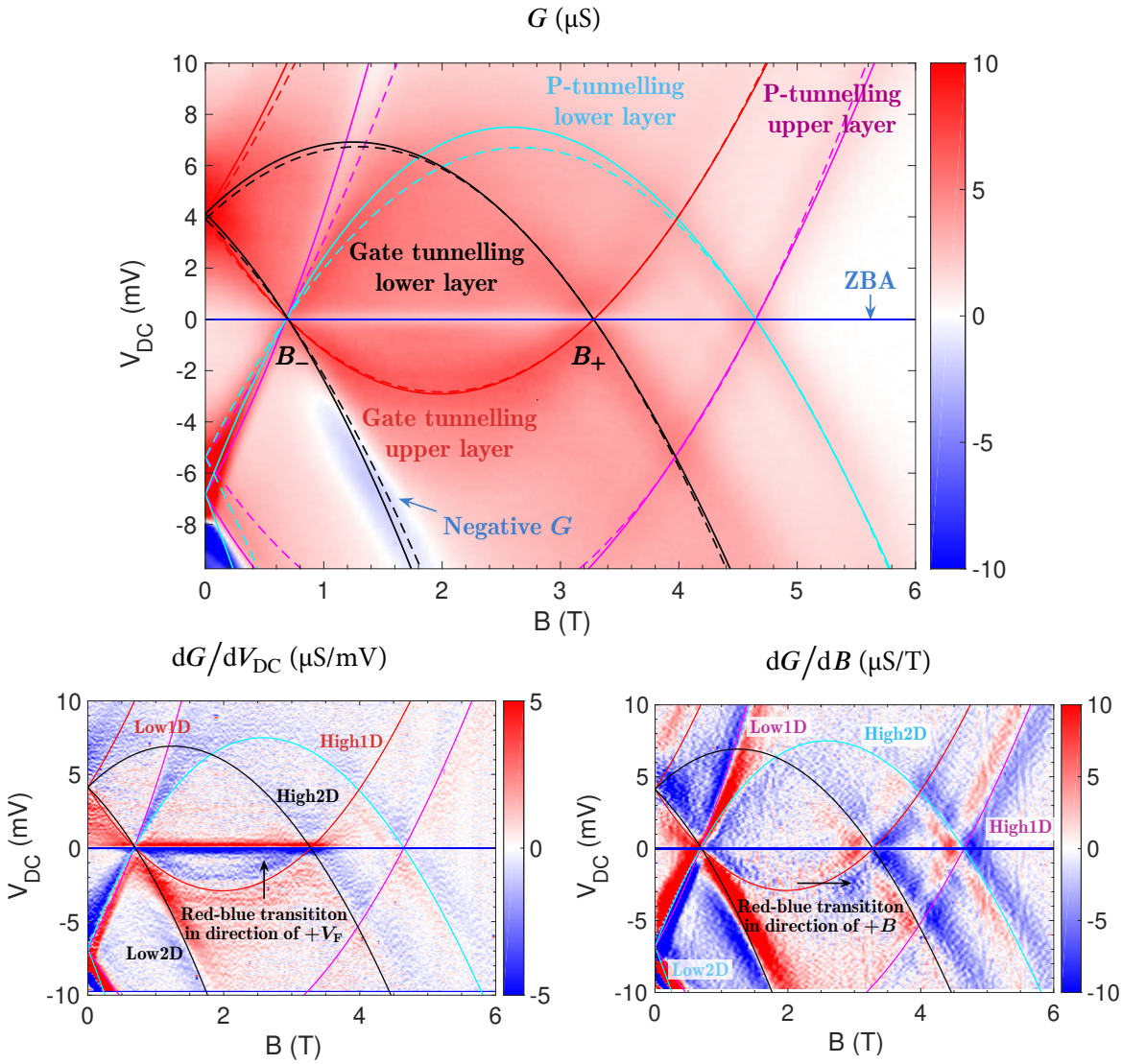
6.5 Analysis of conductance maps

We now make detailed analysis of conductance maps obtained from two separate samples which were measured in two different cryogenic systems. The following sections will demonstrate the process in which the positions of the conductance peaks are determined, as well as how important physical parameters of the sample can be deduced from the process.

6.5.1 Sample I

The first device to be analysed was fabricated on the C2617 wafer (EBL pattern is specified in the CAD file `C2617_03.dwg`). It is device number 4 on die V-I, located closest to the right ohmic contact. The device features a single-column wire-gate array, with individual gates of length 18 μm and width 0.1 μm. The 1D channel width (separation between adjacent wire gates) is 0.18 μm. Since the device contains only a single column of wire gates, air-bridges are not required to connect the repeating units. Instead they are connected at one end by a single metal strip, forming a comb-shaped array. Sample I was measured in a ^{3}He/^{4}He dilution fridge at 50 mK.

Figure 6.9 is a conductance map measured on sample I at wire-gate voltage $V_F = -0.68$ V, where only a single 1D subband is visible. The map contains 200 DC sweeps in total. Note first that the V_{DC} axis has been inverted compared to Figure 6.7. This is done to match the sign convention used in the numerical model, which will be discussed in the next chapter. The inverted sign convention will be used throughout the rest of the thesis unless stated otherwise. Furthermore, a small offset is present on the V_{DC} direction in all the density maps. This is because every measurement was affected by a small DC voltage offset caused by the current pre-amplifier in the circuit. The position of the true zero-DC line can be determined easily by following a horizontal feature on the conductance map known as the zero-bias anomaly (ZBA)—a sharp, B-independent suppression of tunnelling conductance. The anomaly arises from a reduction in the density of states at the Fermi level due to interactions between electrons [46, 22, 81], which has been observed in different systems such as carbon nanotubes [13], tunnelling between quantum wires [4]. The calculation of the density of states in terms of the Luttinger model can be found in [47, 48], and a power-law dependency on temperature is predicted in [31]. The ZBA is marked by the blue horizontal line on the maps. For Sample I, the offset required to align the ZBA line with zero bias was 0.27 mV.



SUBBAND PARAMETERS

n	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	k_{F2D} (m^{-1})	n_{2D} (m^{-2})
1	0.70	3.28	7.06×10^7	4.49×10^7	1.09×10^8	1.89×10^{15}
p	0.68	4.65	1.46×10^8	9.28×10^7	1.09×10^8	1.88×10^{15}

FIGURE 6.9 Conductance map of Sample I and its derivatives in V_{DC} and B, measured at $V_F = -0.68$ V

The conductance peaks are fitted by a series of parabolic curves, calculated as per the discussion in Section 3.2. As for the maps in Figure 6.7, two sets of parabolae can be seen in the conductance map of Figure 6.9. The first set consists of a black and a red parabola, and cor-

responds to the tunnelling signal from the 1D channel array; the second set consists of a cyan and a magenta parabola, and corresponds to the tunnelling signal from the parasitic regions. The red and magenta curves represent the dispersions in the upper layer, while the black and cyan curves are the dispersions in the lower layer. In the terminology of Section 3.2, the dispersions of the upper layer are given by the functions Low/High1D, while the lower layer dispersions are Low/High2D of Equation 3.22. Note that the fitted parabolae are inverted in the V_{DC} direction compared to Figure 3.5, which is caused by the reversed polarity of the DC bias in the experimental setup. Each set of parabolae consists of a dashed and a solid curve, where the dashed one represents the uncorrected dispersions given by Equation 3.22, and the solid one represents their capacitance-corrected counterparts (see the discussion on Equation 3.30). The exact shapes of the curves depend on the positions of crossing points on the zero-DC line ($B_{\pm}$) and the inter-layer capacitance (Γ). In order to fit the dispersion curves to the conductance peaks in the data, we calculate the derivatives of the differential conductance G with respect to V_{DC} and B, which show the peaks more prominently. Conductance maxima appear in the differentiated maps as positions with zero gradients that are flanked by areas of positive-to-negative gradients in the direction of differentiation. Consequently in the symmetric blue-white-red colour scale, the peaks appear as narrow white curves centred between red-to-blue transitions. The fits are determined iteratively, by repeatedly varying the parameters $B_{\pm}$ and Γ until the dispersion curves coincide with the zero derivatives in both differentiated maps. Note that at large distance from the zero-DC line, and at positions where multiple dispersions overlap, it is impossible to match the calculated curves perfectly with the zero gradients. We aim to match the 1D dispersions with the zero gradients below the Fermi energy (zero-DC line) because the tunnel signals here are strongest and the DC bias is low (therefore less susceptible to capacitance effects). Notably we found it necessary to adjust the inter-layer separation d to make the curves match the data: instead of using the nominal value of 32 nm from the MBE growth data, the data is better fitted with $d = 36\,\mathrm{nm}$. This may be explained as a result of the electron wavefunctions being further apart than the centre-to-centre distance between the two quantum wells, due to mutual repulsion. We also found that although the capacitive correction makes better fits to data at low DC levels ($\left|V_{\mathrm{DC}}\right| < 5\,\mathrm{mV}$), it cannot account for the deviation of data at high DC bias ($\left|V_{\mathrm{DC}}\right| \gg 5\,\mathrm{mV}$). However, as the later fitting work on the non-linear TLL model will be carried out close to the 'trough' of the upper-layer dispersion, which is above −5 mV, we do not pursue further correction to the dispersion formula in this work. For the dispersion fits in Figure 6.9, the parameters of the wire-gate tunnelling signal are $B_{-} = 0.70\,\mathrm{T}$, $B_{+} = 3.28\,\mathrm{T}$, and an inter-layer capacitance of $\Gamma/A = 0.0012\,\mathrm{F\,m^{-2}}$; the parasitic region's parameters are: $B_{-} = 0.68\,\mathrm{T}$, $B_{+} = 4.65\,\mathrm{T}$, and $\Gamma/A = 0.0047\,\mathrm{F\,m^{-2}}$. Parameters such as the Fermi wavevector,

electron density and Fermi energy can be calculated from the crossing points using Equation 3.24 and 3.25. See the table at the bottom of the figure for the results. From Equation 3.31, the per-unit-length (m^{-1}) values of the inter-layer capacitance are $\Gamma/L = 68.0$ pF in the channels and 129 pF in the p-region.

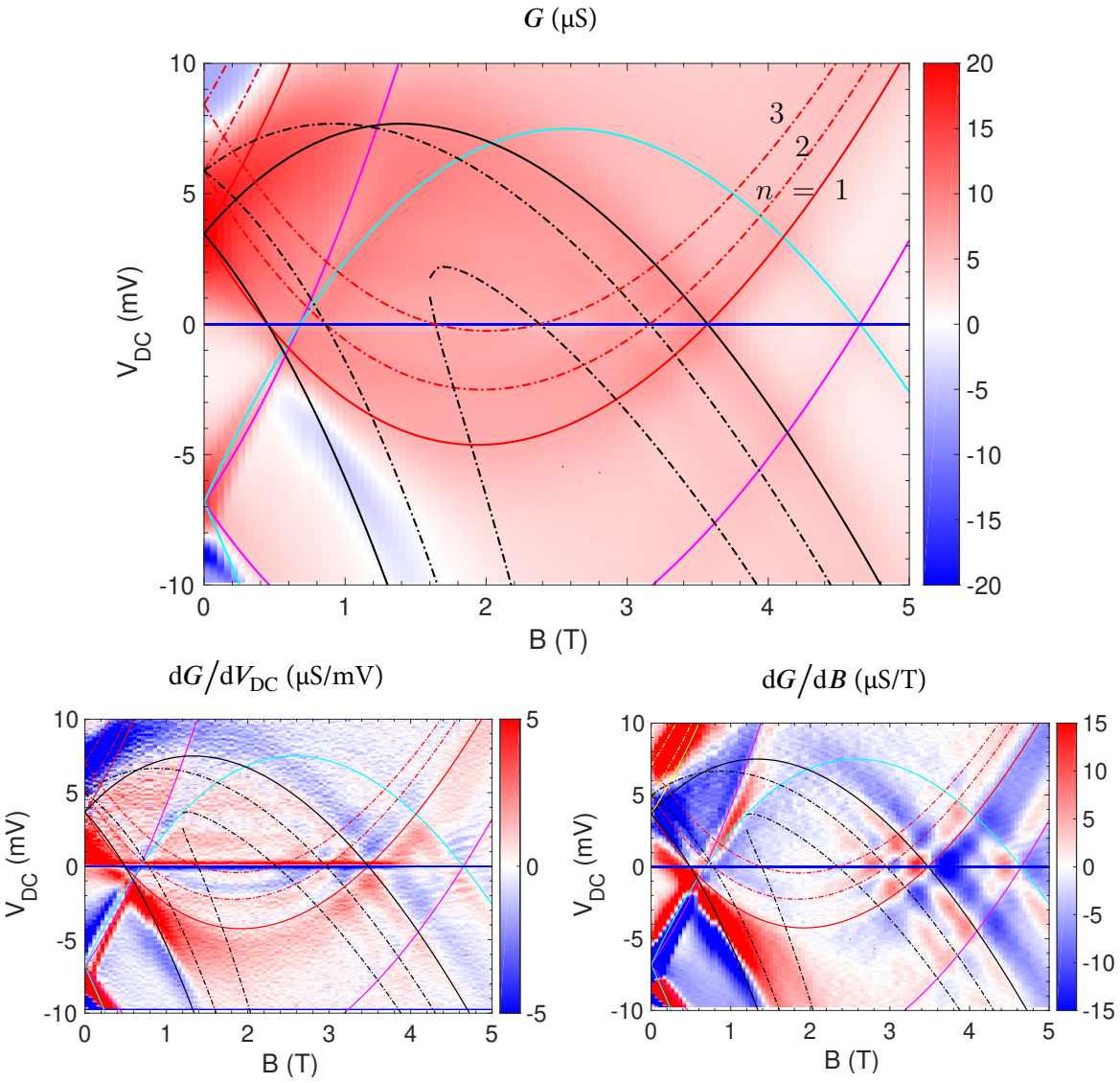
Having analysed a result in the single-subband regime, we can now study more complicated results with more than one 1D subband. Figure 6.10 shows a conductance map obtained from Sample I at wire gate voltage −0.55 V. It consists of 100 DC sweeps between the B range 0 to 5 T, slightly shorter than the previous result. The DC offset required to zero the ZBA line is 0.27 mV. The dispersions are marked by parabolae in the same line styles as the previous analysis, although in some parts of the map, 1D dispersions are plotted in yellow instead of red to better distinguish them from the background. Three dispersions can be resolved on this map, indicating that three 1D subbands lay below the Fermi energy at this voltage. The 1D subbands are labelled by the index $n = 1$, 2 and 3. The first and second 1D subbands are represented by solid and dot-dashed curves respectively. They are the first to be determined in the analysis process, by altering $B_\pm$ and Γ iteratively until the curves best match the zero gradients with the portions of the subbands that are below the zero-DC line, which represents the Fermi energy. The fitting of the third subband is carried out with a different method: because it is almost empty, the portion of the dispersion below the zero-DC line is too small to be used as a fitting reference. Instead, we estimate its position by noting that the centres of each 1D dispersion should line up approximately in the middle—we may assume the 1D channels have parabolic confinement potentials, which implies that the subband dispersions would line up as the ladder eigenvalues of harmonic oscillators. Denoting the crossing points between the third subband and the B axis ($V_{\mathrm{DC}} = 0$) as $B_{\pm,n=3}$, the alignment of the subbands can be expressed as

$$\frac{B_{+,3} + B_{-,3}}{2} = \frac{B_{+,2} + B_{-,2}}{2}, \tag{6.1}$$

where we have equated the centre of the third subband with that of the second subband. Let the crossing point between the n-th subband and the V_{DC} axis ($B = 0$) be $V_{0,n}$. Equation 3.22 shows that at $B = 0$, the subband dispersions are:

$$eV_{0,n} = \frac{\hbar^2}{2m^*} \cdot \begin{cases} k_{\mathrm{F1},n}^2 - k_{\mathrm{F2}}^2 & \text{Lower layer} \\ k_{\mathrm{F2}}^2 - k_{\mathrm{F1},n}^2 & \text{Upper layer} \end{cases} \quad \Longrightarrow \quad V_{0,n} = \frac{ed^2}{2m^*} \cdot \begin{cases} B_{+,n}B_{-,n} \\ -B_{+,n}B_{-,n} \end{cases}, \tag{6.2}$$

where $k_{\mathrm{F1},n}$ is the Fermi wavevector of the n-th subband. The second equality is obtained by converting $k_{\mathrm{F},n}$ into $B_{\pm,n}$ using Equation 3.23. Combining Equations 6.1 and 6.2 yields a



SUBBAND PARAMETERS

n	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	k_{2D} (m^{-1})	n_{2D} (m^{-2})
1	0.49	3.48	8.18×10^7	5.21×10^7	1.09×10^8	1.88×10^{15}
2	0.78	2.96	5.96×10^7	3.80×10^7	1.02×10^8	1.66×10^{15}
3	1.39	2.35	2.62×10^7	1.67×10^7	1.02×10^7	1.66×10^{15}
p	0.68	4.65	1.46×10^8	9.28×10^7	1.09×10^8	1.88×10^{15}

FIGURE 6.10 Conductance map and its derivatives in V_{DC} and B of sample I, measured at $V_F = -0.55$ V.

relationship between $B_{\pm,n=3}$ and $V_{0,n=3}$:

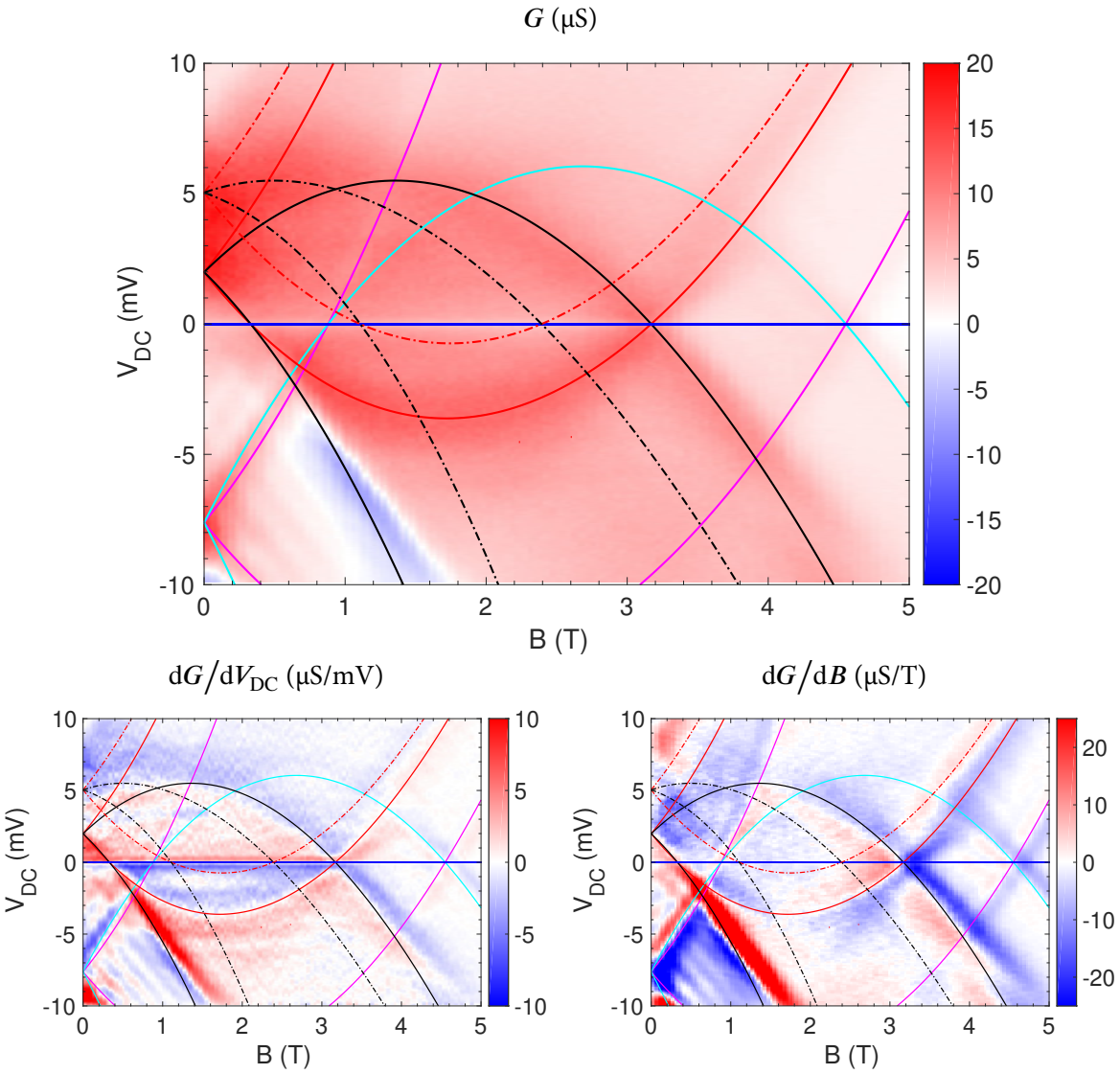
$$\left(B_{+,3} - B_{-,3}\right)^2 = \left(B_{+,3} + B_{-,3}\right)^2 - 4B_{+,3}B_{-,3}$$

$$\implies \quad \frac{B_{+,3} - B_{-,3}}{2} = \sqrt{\left(\frac{B_{+,2} + B_{-,2}}{2}\right)^2 \pm V_{0,3}\frac{2m}{ed^2}} \tag{6.3}$$

where the plus/minus sign inside the square root applies to the dispersions of the upper/lower layer, respectively. Using this equation, we can determine the position of the last subband by varying a single parameter $V_{0,3}$ and plotting the dispersions with the resultant $B_{\pm,n}$ values until the plots match the data. The signal of the third subband is very weak relative to the other two subbands because it is close to the Fermi level. Even though it is almost empty, the electrons in the 2D layer can still tunnel into it (as long as there are empty states and $V_{\mathrm{DC}} > 0$), and the effect of this can be seen in the top left corner of the map, around 5 to 10 mV near $B = 0$. Conversely, tunnelling effects from the lower layer into the upper layer can be seen between −5 to −10 mV and 3 to 4 T. These conductance signals can be used to fix the position of the last dispersion as we vary the fit parameter $V_{0,3}$. The parameters calculated from the crossing points are summarised in the table in Figure 6.10. The average subband separation is 1.90 meV.

6.5.2 Sample II

The second device to be analysed was fabricated on the W938 wafer. Its EBL pattern is defined in the CAD file `W938_W939_S2.dwg`. It is device number 1 on die V-V, located closest to the left ohmic contact. The wire-gate array on this device consists of three columns of repeating units, interlinked by air-bridges on both ends of the channel. The wire gates measure 10 µm in length and 0.3 µm in width, and the 1D channel width is 0.2 µm. The device was measured in a sorption-pumped ^{3}He cryostat at temperature 300 mK. Figure 6.11 shows the conductance map of Sample II, as well as its derivatives along the V_{DC} and B axes, measured at wire-gate voltage −0.7 V. Two 1D dispersions can be resolved in the figure, which are marked by parabolic curves in the same scheme as before. The second subband is close to the Fermi level and is hence fitted using the same technique used to fit the last 1D subband in the previous section. The DC offset required to align the ZBA is 0.11 mV. The inter-layer distance d which produces the best fits is 36 nm. The inter-layer capacitance in the 1D channel region is $\Gamma/A = 0.0024\ \mathrm{F\,m^{-2}}$, and $0.0047\ \mathrm{F\,m^{-2}}$ in the p-region. The crossing points and the dispersion parameters are listed in the table. The values of the capacitance correction are $\Gamma/L = 123\ \mathrm{pF\,m^{-1}}$ in the 1D channels and $127\ \mathrm{pF\,m^{-1}}$ in the p-region. Note that a series of fringe-like interference patterns can be seen below the black parabola at $B < 1.4$ T. This



SUBBAND PARAMETERS

n	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	k_{2D} (m^{-1})	n_{2D} (m^{-2})
1	0.33	3.17	7.77×10^7	4.94×10^7	9.57×10^7	1.46×10^{15}
2	1.11	2.39	3.52×10^7	2.24×10^7	9.57×10^7	1.46×10^{15}
p	0.88	4.55	1.48×10^8	9.45×10^7	1.00×10^8	1.60×10^{15}

FIGURE 6.11 Conductance map and its derivatives in V_{DC} and B of sample II, measured at $V_F = -0.7$ V.

feature is particularly prominent in the results obtained from Sample II. It is believed to be related to the multi-column geometry of the wire gate array of this device, and is more pro-

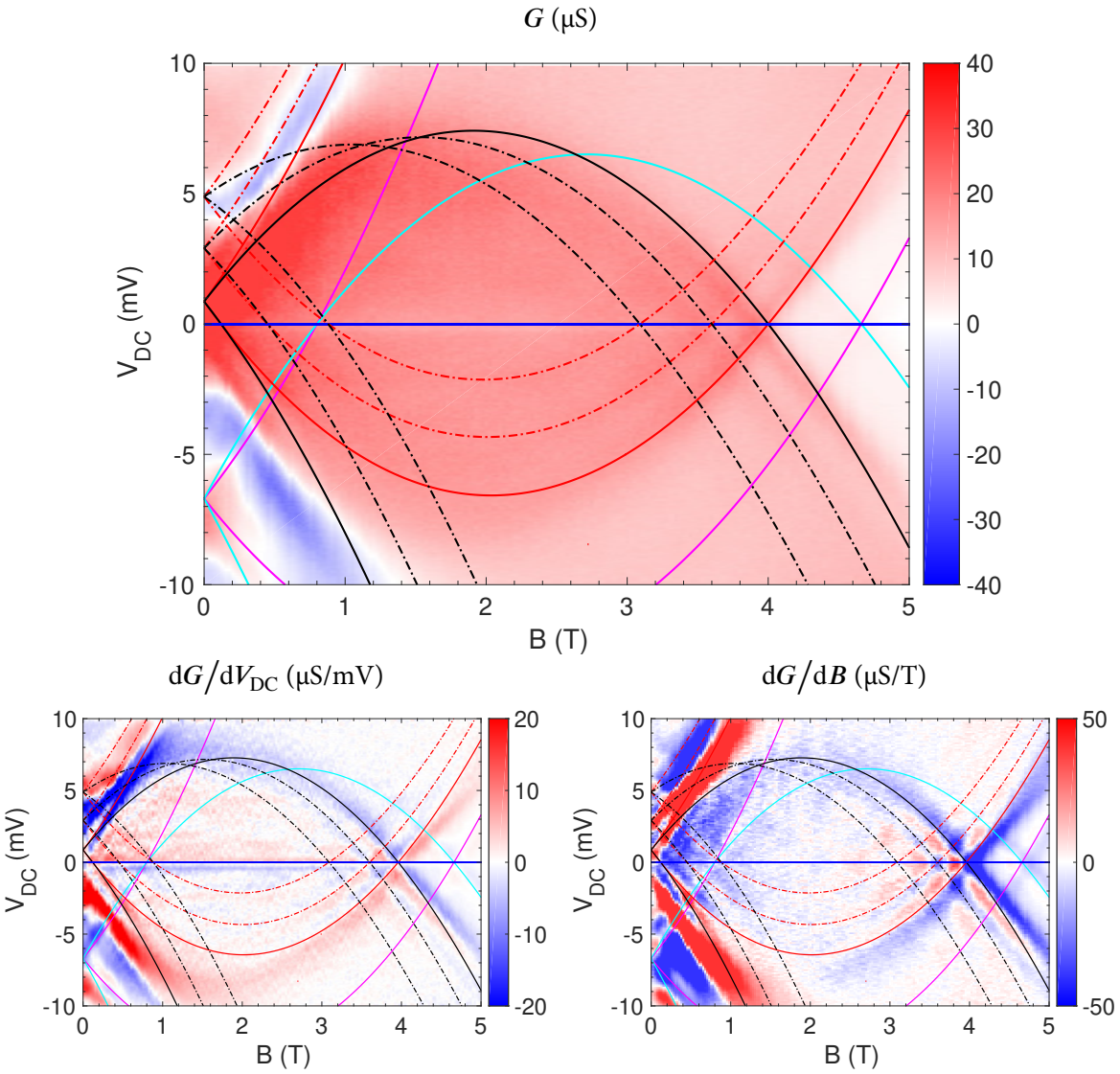
nounced with larger numbers of columns of repeating units [64]. The origin of these patterns is not fully determined. However, they are related to the tunnelling in the p-region and are postulated to be dependent on the periodicity of the wire gate columns.

Figure 6.12 shows the conductance map measured on Sample II at wire gate voltage −0.5 V, consisting of 200 V_{DC} sweeps and the ZBA was zeroed with a DC bias of 0.11 mV. Due to the noise in measurement, the $\boldsymbol{B}$-differential was treated with additional smoothing processes of binning (every three data points) and running average (every five sweeps). Three 1D subbands can be resolved in this result–see the table at the end of the figure for the dispersion parameters. The average subband separation is 2.2 meV. Unlike the results of Sample I presented in the last section, the p-region crossing points are slightly varied between the results obtained from Sample II. This can be attributed to the different experimental procedure: the carrier density of a sample varies slightly depending on the cool-down process. Sample I was measured in a dilution fridge which was capable of continuous operation, whereas Sample II was measured in a sorption-pumped cryostat, which only provides single-shot operations. The Sample I results have identical p-region parameters because they were measured during the same cool-down, whereas Sample II results were obtained in different runs.

6.6 Summary

This chapter gives an overview on the setup and observations of the experimental aspect of the project. At the start of the chapter, we discussed the device design and looked at the function of each lithographic feature of the device. We saw that crucially, the operation of the samples replied on the formation of independent ohmic contacts to the two layers of electrons in the device. Experimentally, this required us to verify the functionality of the split/mid-gates. Testing the split/mid-gates involved verifying quantised conductance through the sample as the mid-gates are squeezed by negative bias. By varying the mid-gate bias as the split-gates are depleted, one can control the layer(s) of 2DEG through which current can travel. For independent ohmic contacts, we require the split/mid-gates to create a conducting channel only to the upper layer. Once this is achieved, we can look for tunnelling conductance by sweeping the bias of the wire-gate array. One last test before detailed measurement commenced was the uniformity of the wire-gate array. This was done by looking at the zero-magnetic field conductance data. In particular, we verify that the data is able to resolve 1D subbands in the upper layer.

In the second half of the chapter, we discussed the details of the two-terminal measurement, the main method of obtaining tunnelling conductance data from our device. Results obtained from this experiment are intensity maps of the tunnelling conductance with the



SUBBAND PARAMETERS

n	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	k_{2D} (m^{-1})	n_{2D} (m^{-2})
1	0.12	3.96	1.05×10^8	6.69×10^7	1.12×10^8	1.98×10^{15}
2	0.45	3.60	8.61×10^7	5.48×10^7	1.11×10^8	1.95×10^{15}
3	0.88	3.09	6.04×10^7	3.85×10^7	1.09×10^8	1.88×10^{15}
p	0.80	4.66	1.49×10^8	9.50×10^7	1.06×10^8	1.77×10^{15}

;q

FIGURE 6.12 Conductance map and its derivatives in V_{DC} and B of sample II, measured at $V_F = -0.5$ V.

magnetic field and DC-bias as independent variables. The most important analysis on the conductance data is determining the electronic dispersions of the two layers. For this we refer to the theory of resonant tunnelling that was detailed in Chapter 3. Following the theoretical discussion, we expect the dispersions to follow the maximum tunnelling conductance. Using the conductance map, we calculated the differential conductance in the $\boldsymbol{B}$ and DC-bias directions, then plotted capacitance-corrected dispersion parabolae which best fitted the observed conductance maxima. From the zero-DC crossing points of the best-fit dispersions, we can calculate the Fermi wavevectors and electron densities of each subband in the data. Having calculated the positions of the dispersions, we are now ready to study the conductance data in further detail.

Chapter 7

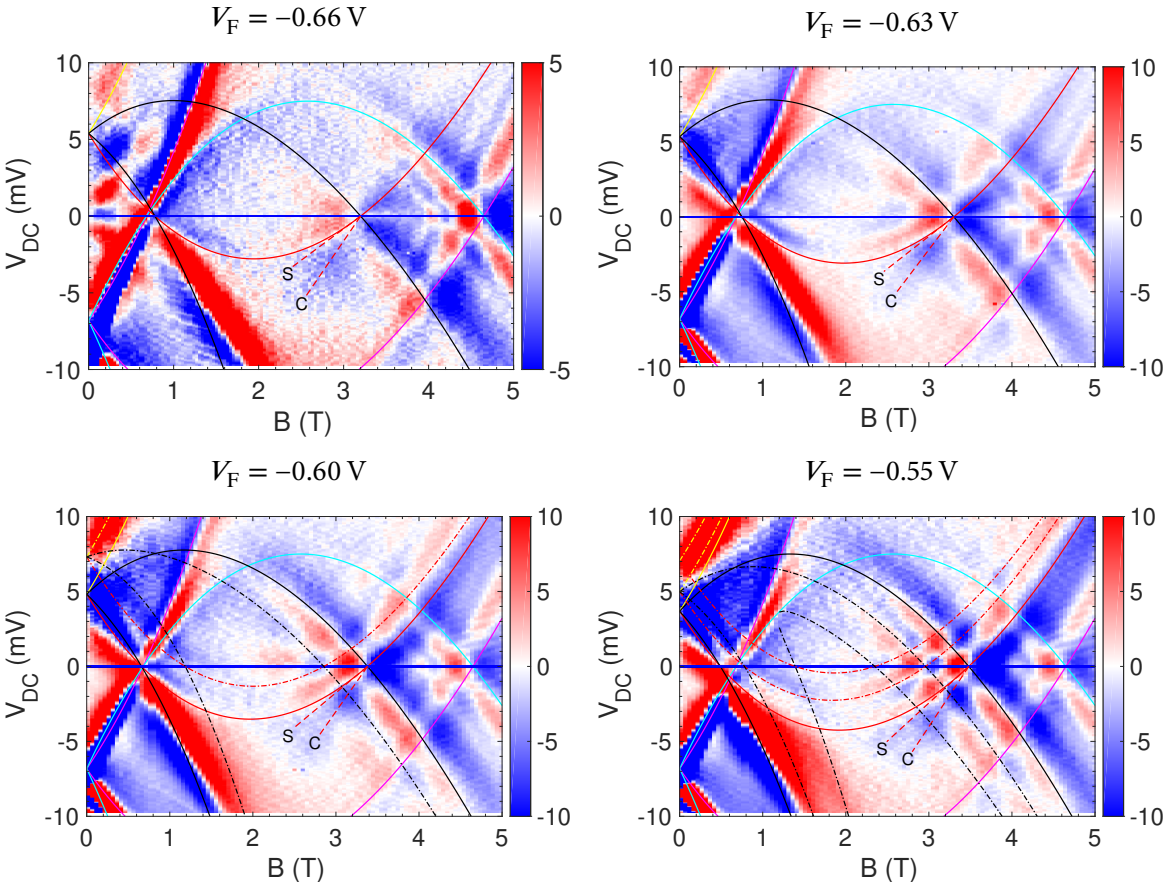
Interaction effects in tunnelling data

Analysis in the previous chapter has demonstrated that non-interacting single-particle electron dispersion can be used to fit the positions of the maximum tunnelling conductance with good agreement at low energies. In this chapter, we will further analyse the conductance data and identify features on the conductance map that lie outside the non-interacting fits. In particular, we will search for effects related to electron-electron interactions, and determine whether the observed interacting effects are in agreement with the prediction of the theory outlined in Chapter 1.

7.1 Spin-charge separation

In Section 1.3.2, it was shown that a hallmark property of the spinful Luttinger-liquid model is that the excitations in an 1D interacting fermionic system are hydrodynamic-like modes, which split into charge- and spin-density waves. For finite interaction strengths, the two types of modes have linear dispersions with distinct slopes v_c and v_s, which correspond to their respective velocities. In the absence of interactions, the difference between the two velocities vanishes, and the spin degree of freedom only results in the double degeneracy observed in the free-electron model. For repulsive interactions such as the Coulomb interaction between electrons, it is expected that $v_c \geq v_s$. The ratio v_c/v_s can therefore be used as a parameter which characterises the interaction strength, where higher ratio indicates stronger interaction. The divergence of the two velocities has been observed experimentally in 1D-1D tunnelling between closely-spaced parallel quantum wires [4], as well as in 1D-2D tunnelling as was reported by the precursor project to this work [45]. [32, 44] presented a numerical evaluation of the tunnelling current from a model TLL spectral function broadened by physical effects. By analysing the calculated conductance, the authors concluded that spin-charge separation manifests itself in the 1D-2D tunnelling conductance map as a conductance peak

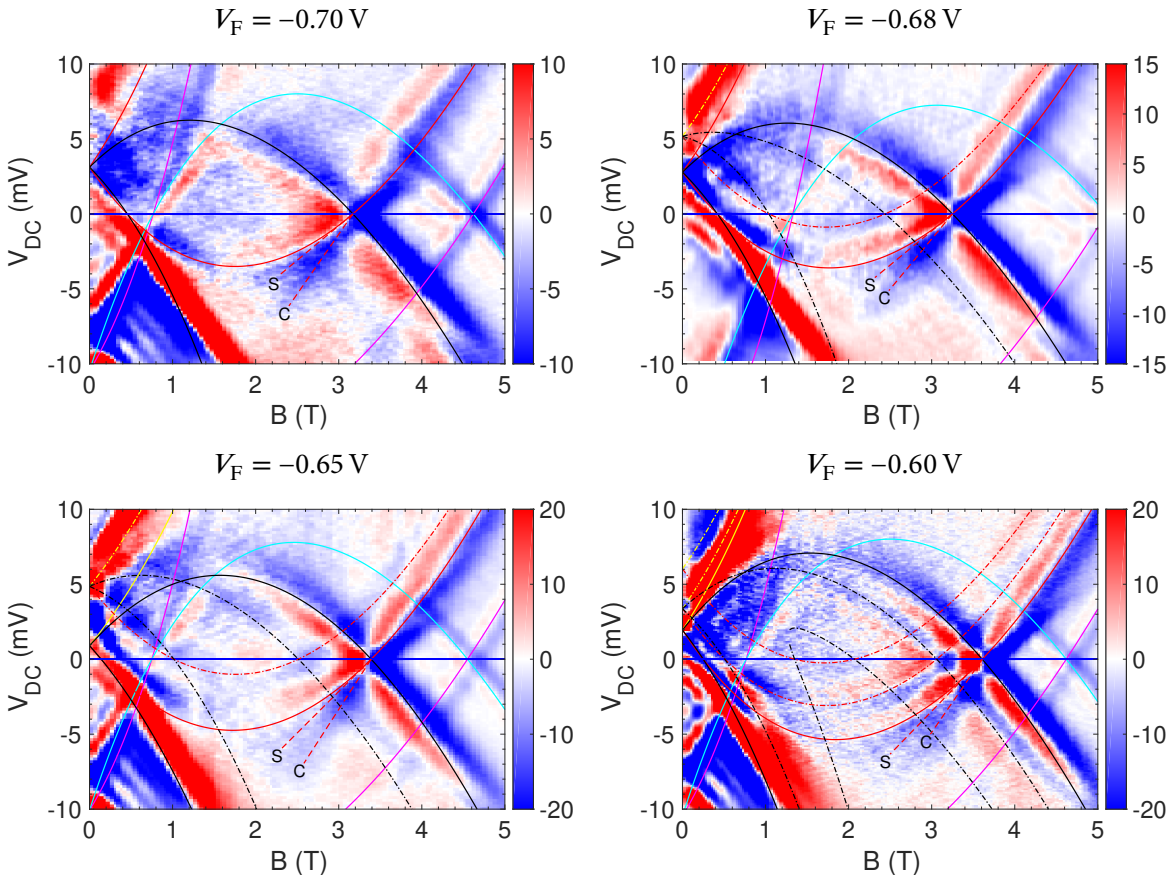
(attributed to the spin mode of excitation) and a sharp drop of conductance away from the peak (attributed to the charge mode). They suggested that this effect is observed most easily in the magnetic-field differential $\mathrm{d}G/\mathrm{d}B$, where the spin mode is found by following the tangent of the non-interacting dispersion, and the charge mode by finding the positions with the most negative gradient (deepest drop in conductance). We adopt the method of the referenced works to data gathered in our experiment.



PARAMETERS OF 1ST 1D SUBBAND

V_F (V)	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	n_{2D} (m^{-2})	v_c/v_s
-0.66	0.78	3.20	6.62×10^7	4.21×10^7	1.89×10^{15}	2.10
-0.63	0.75	3.30	6.97×10^7	4.44×10^7	1.95×10^{15}	1.69
-0.60	0.66	3.38	7.44×10^7	4.74×10^7	1.94×10^{15}	1.71
-0.55	0.49	3.48	8.18×10^7	5.21×10^7	1.88×10^{15}	1.65

FIGURE 7.1 Density maps of $\mathrm{d}G/\mathrm{d}B$ of Sample I showing signatures of diverging features of spin and charge mode excitations (spin-charge separation). The gradients of the dashed lines labelled as S (spin) and C (charge) correspond to the velocities of the two types of density waves.



PARAMETERS OF 1ST 1D SUBBAND

V_F (V)	B_- (T)	B_+ (T)	k_{F1D} (m^{-1})	n_{1D} (m^{-1})	n_{2D} (m^{-2})	v_c/v_s
-0.70	0.46	3.17	7.42×10^7	4.73×10^7	1.56×10^{15}	1.73
-0.68	0.42	3.25	7.52×10^7	4.79×10^7	1.52×10^{15}	1.47
-0.65	0.14	3.39	8.89×10^7	5.66×10^7	1.48×10^{15}	1.55
-0.60	0.25	3.61	9.18×10^7	5.85×10^7	1.77×10^{15}	1.41

FIGURE 7.2 Density maps of dG/dB of Sample II showing signatures of spin-charge separation.

Figure 7.1 shows the magnetic-field differentials of the tunnelling conductance of Sample I at four different wire-gate voltages. The spin and charge mode gradients are highlighted on the maps by two diverging red dashed lines near the high field crossing point of the first 1D dispersion. The colour scales of the maps have been chosen to enhance the conductance near B_+, which makes the rest of the maps appear over-saturated. The spin-wave velocity is assumed to be the same as a non-interacting electron's, hence it is given by the gradient of the non-interacting dispersion at $V_{DC} = 0$. Following the 1D dispersions of the first subband (solid red curves) near the high-field crossing point, we see that while the positive gradient

(on top of the curve) follows the parabolic curvature of the dispersions closely, the negative gradients (below the curves) diverge from the curvature. The charge-mode line is taken to be in the middle of the diverging negative gradient. The table at the bottom of Figure 7.1 summarises the parameters of the fits to the first 1D subbands. v_c/v_s lies in the range 1.6 to 1.7 for experiments in the wire gate voltage range −0.55 to −0.63 V, and rises quickly as the channels approach pinch-off (which occurs at ∼−0.7 V). The presence of the second 1D subband changes the screening radius for the Coulomb interaction potential in the first subband significantly, which is reflected in the smaller value of v_c/v_s when multiple 1D subbands appear below the Fermi energy. Figure 7.2 shows a set of similar data obtained from Sample II, which has a v_c/v_s which varies between 1.4 to 1.5, and rises past 1.7 when the wire gates are biased towards pinch-off—a behaviour similar to that observed previously.

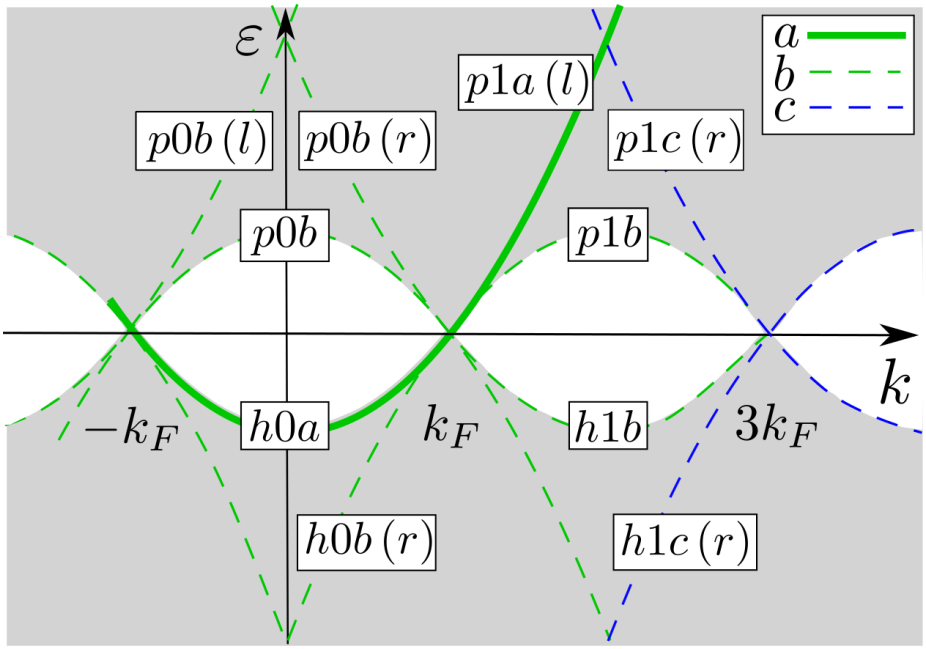


FIGURE 7.3 Main features of the spectral function from Bethe Ansatz calculation. The white region marks the area where the spectral function vanishes, and the grey region marks where it is non-zero. Curves in style *a*, *b* and *c* represent the hierarchies of order 0, 1 and 2, respectively. The labels identify each unique branch—see the referenced article for more details on the labels. Figure taken from [93].

7.2 Hierarchy of excitation modes

The spin-charge separation effect shown in the last section is a well-known prediction of the linear Luttinger-liquid model, which has limited range of validity at low-energy excita-

tions. A recent theoretical method for solving the 1D fermionic many-body problem generally, without the linearity assumption is put forth by Tsyplyatyev et al. [93, 92], which involves the full microscopic diagonalisation of the spinless one-dimensional fermion Hamiltonian using the Bethe Ansatz. The solution obtained by this method consists of a "hierarchy" of spectral functions, characterised by a factor of integer powers of a length ratio $\mathcal{R}^2/L^2$, where $\mathcal{R}$ is the interaction radius (taken to be the Fermi wavelength) and L the length scale of the system (taken to be the length of wire gates). Figure 7.3 depicts the main features of the spectral function. The first-level excitation forms a parabola which resembles single-particle dispersion, although it has a mass renormalised by the Luttinger parameter K [91]. The second- and third-level excitations consist of branches of replicas of the first-level mode with higher-order prefactors. The exact forms of the spectral functions can be found in the referenced articles. Experimentally, we observe distinctive features in the conductance maps around the high-field crossing point which resemble the $p1b$ mode: Figure 7.4 shows three density maps of the V_{DC} differential of conductance around the vicinity of B_+. A feature can be seen in all three maps that diverges away from the first subband dispersion (solid red curve), and closely follows the calculated curvature of the second-level excitation $p1b$ mode (dashed green curve). Indeed, this second-level feature can be seen in most results already shown in the previous sections, most notably in the magnetic-field differential maps, even in the presence of multiple 1D subbands. For simplicity, we limit our analysis to the results with single 1D subband. The line plots which accompany each density map consist of cuts (non-consecutive) of conductance in the V_{DC} direction between 3.5 to 4.5 T. The '+' labels highlight the voltage values along the second-level excitation (dashed curve), and the '×' labels highlight the voltage along the first 1D subband. The line plots show that the conductance along the second-level mode drops off at a fast rate with increasing B, which is analogous to the power law calculated for spinless fermions in [92]. In terms of the results obtained from Sample I (first two rows of results in Figure 7.4), the typical Fermi wavevector in the single-subband regime is $\sim 7 \times 10^7$ m^{-1}, which is equivalent to a Fermi wavelength of ∼90 nm. Using the nominal length of the wire gates (10 µm) as the length scale L, one expects the second-level excitations to be smaller by a factor of $\lambda_{\mathrm{F}}^2/L^2 \approx 8 \times 10^{-5}$. Since the conductance of the principal parabola is ∼6 µS, the expected conductance due to the second-level excitation ($\sim 10^{-3}$ µS) is at least one order of magnitude smaller than the noise levels of our experiment ($\sim 10^{-2}$ µS). This is consistent with the observed picture of rapidly vanishing second-level effects. The measurement of a separate device with 1 µm-long wire gates is reported in [64], where the full second-level excitation is visible.

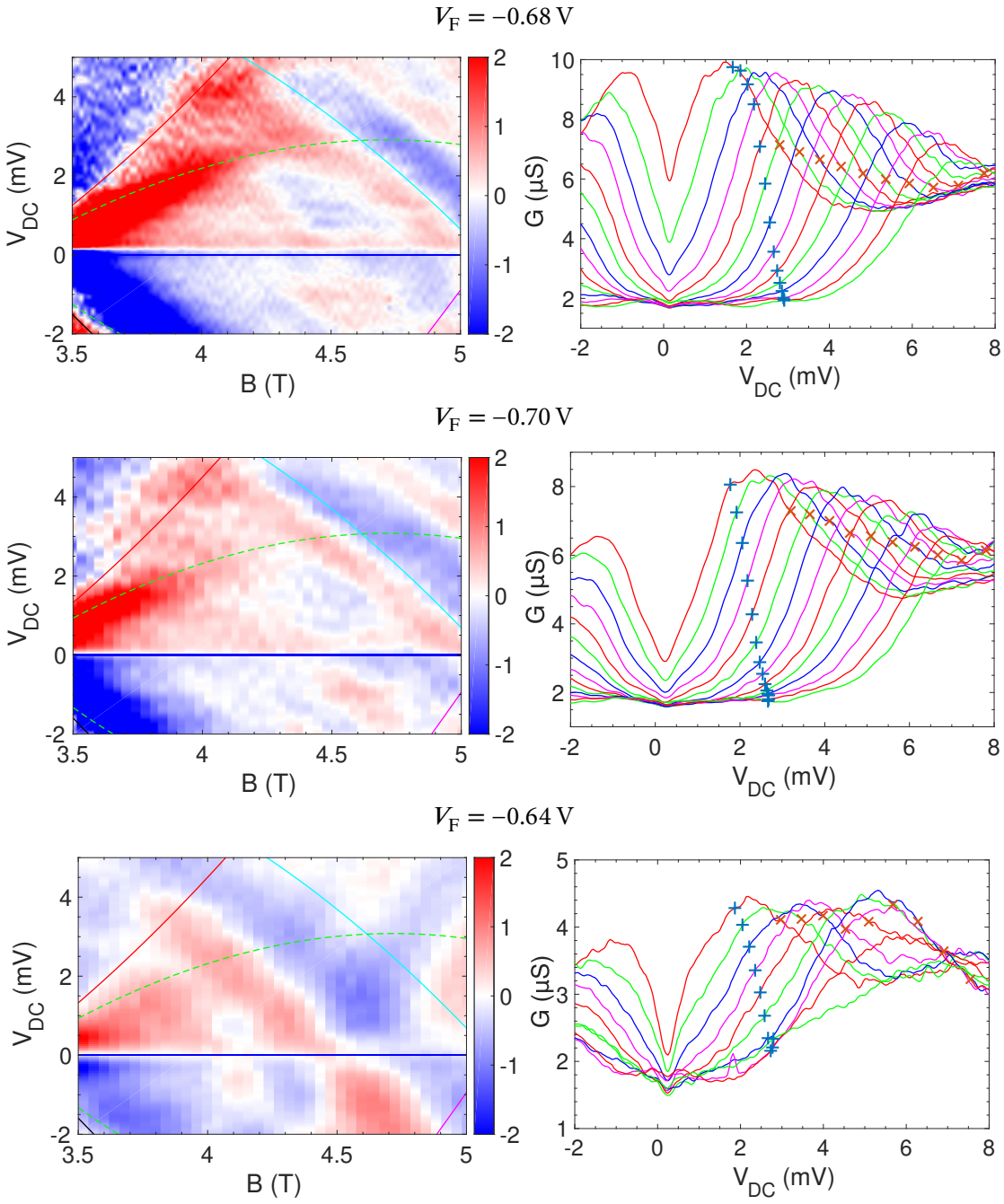


FIGURE 7.4 Density maps of $\mathrm{d}G/\mathrm{d}V_{\mathrm{DC}}$ around the high-field crossing point B_+, and cuts of the maps in the V_{DC} direction. The first two rows of results were measured from Sample II, and the last row of result was from Sample I. The line cuts were taken (non-consecutively but with equal separation) between 3.5 to 4.5 T, with the '+' symbol showing the voltage along the first-level excitation (green dashed curve in the left plot), and the '×' symbol the first 1D subband (red solid curve to the left). The results were taken in the single-1D-subband regime, showing the conductance feature which diverges away from the main 1D dispersion, and follows the second-level excitation calculated from the Bethe-Ansatz method.

7.3 Momentum-dependent power law

As was introduced in section 1.4.1, the mobile-impurity model is a phenomenological, non-linear model which extends the low-energy universality of the linear Luttinger-liquid model to finite energy. The key result of the discussion in the first chapter is that the model is a hydrodynamic theory similar to the TLL, but takes into account the finite curvature of the dispersion which changes with momentum. Consequently, it leads to a power-law spectral function (Equation 1.113), which in contrast to that of the linear TLL model (Equation 1.105), acquires a momentum dependency in the power-law exponent (Equation 1.117).

A method for numerically evaluating the tunnelling current and conductance between two layers of 2DEG was discussed in Section 3.1.1, which involves the evaluation of Equations 3.12–3.15. Having analysed the data gathered the tunnelling experiment, we are in a position to compare the measurement against calculations based on the different theoretical models. The following analysis was first reported in [43]. Three calculations are carried out, where the spectral function of the upper layer will be separately represented by the results from the non-interacting model (Equation 3.18), the constant power-law model of the linear TLL theory (Equation 3.19) and the momentum-dependent power-law model of the non-linear theory (Equation 3.21). Each calculation requires two physical parameters to be specified: the value of the spread of the spectral functions Γ, and the value of the interaction parameter K_c. We do not specify K_s, because in spin-invariant systems $K_s = 1$. Additionally, a prefactor is added to the electron effective mass in the model as the third parameter. The conductance is calculated numerically using Mathematica (Calculation details can be found in the Mathematica files `CalcTunnelG_v10_1Dmasschange.nb` and `CalcTunnelG_Formula_v10_1Dmasschange.nb`). Since the computation of the scripts was very time-consuming, instead of fitting the model with the data using an algorithmic routine, we opted to calculate the model using a fixed list of input parameters and visually compare the results with the data.

Before the data can be compared against the calculations, it must be treated so that the conductance signal from outside the 1D-2D tunnelling region is filtered out as much as possible. As the main source of extra conductance comes from the p-region, the filtering is achieved by subtracting the target data with another set of data that has the same p-parabolae crossing points and was measured with the wire gates at pinch-off, using identical sweep parameters. Figure 7.5 and 7.6 are two sets of conductance density maps measured on Sample I and II, respectively. The original data and the background conductance used to generate the filtered map are displayed below the main figures, together with the differentiated maps of the conductance after background-elimination. Note that although a second subband is visible at $\sim$−2 mV in Figure 7.5, its parabola is not displayed, because the effects due to sec-

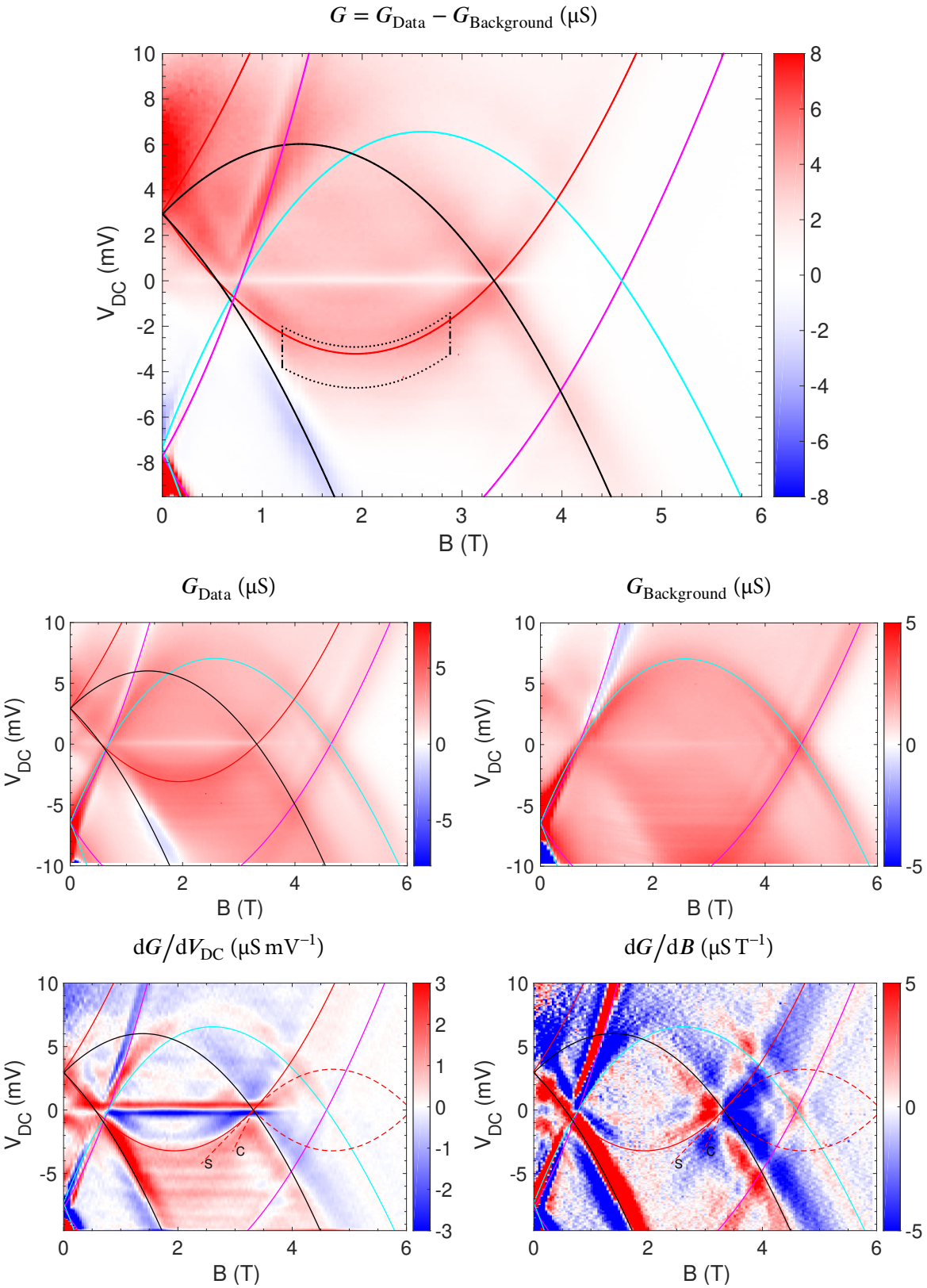


FIGURE 7.5 Density maps of the tunnelling conductance of Sample I, measured at wire gate voltage $V_F = -0.63$ V. G is the difference between the raw data (G_{Data}, measured at $V_F = -0.63$ V) and the background ($G_{\text{Background}}$, measured at $V_F = -0.70$ V). The dotted area around the 1D dispersion in the first figure shows the region of validity of the mobile-impurity model, where the data is compared to the calculations of the models.

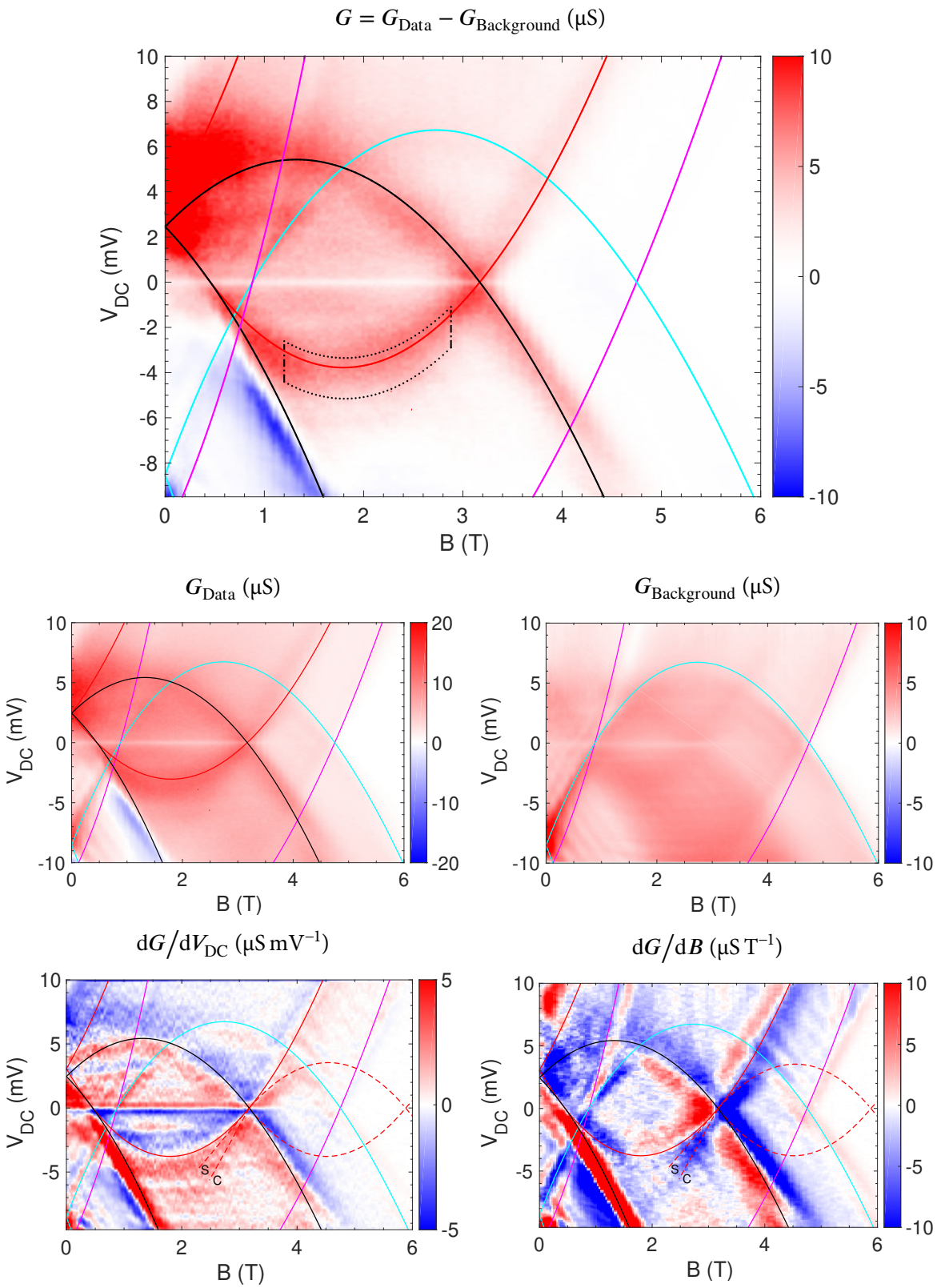


FIGURE 7.6 Density maps of the tunnelling conductance of Sample II, measured at wire-gate voltage $V_F = -0.70$ V. G is the difference between the raw data (G_{Data}, measured at $V_F = -0.70$ V) and the background ($G_{\text{Background}}$, measured at $V_F = -0.75$ V).

ond (and higher) subbands are ignored in the model calculation. On the differentiated maps, the spin/charges lines and the replica-mode lines are displayed, demonstrating the presence of strong interaction effects during device operation. An area following the curvature of the principal 1D parabola is highlighted by black dotted lines in the main figures, showing the region where comparison will be made between the data and the model calculations. Note this region extends past the spectral edge to be slightly above, even though the spectral function of the 1D layer is zero above the spectral edge (Equation 1.113)—this is because the conductance is given by the convolution of the spectral functions of the upper and lower layers. The data is taken as cuts through the main conductance map along the V_{DC} axis between the ZBA line and −4 mV. The B values at which the cuts are made are chosen to be far between the spin/charge lines around B_+ and the Low2D parabola, so that these features do not overlap with the measured conductance.

Figure 7.7 shows the tunnelling conductance calculated with the three different models at magnetic field $B = 2.62$ T, a typical value between B_- and B_+ in our experiment. The calculations were performed in the V_{DC} range (−5 to 1 mV) where the 1D dispersion peak is expected to appear. Each model has been normalised with respect to its peak conductance. The interacting model has two parameters: the interaction parameter K_c and a Lorentzian spread Γ_{Int}; the non-interacting model has a single parameter: the Lorentzian spread Γ_{NonInt}, and the constant power-law model has two parameters: the power-law exponent α and the spread Γ_α. Subfigure (a) shows the three models at the same spread Γ =0.3 meV, our lower-bound estimate on the data's spread. Note that the interacting model predicts an enhancement in conductance to both side of the peak compared to the non-interacting model. Meanwhile the constant power-law model has an asymmetric peak with positive skewness at $\alpha = 0.86$. Subfigure (b) shows the models at the same Γ (0.3 meV) but with increased K_c for the interacting model and α for the constant power-law model. Using the non-interacting model as the reference (as it is identical to Subfigure (a)), we see that the effect of increasing K_c and α is a negative skew in the respective peaks. Additionally, the constant power-law model approaches the non-interacting model as α approaches 1, and overlaps with it completely at $\alpha = 1$, as expected from theory. In subfigure (c) and (d), we have superimposed a data series measured from the wafer C2617 sample at 50 mK on Fridge 2. The data series has a background conductance subtracted and is normalised to its peak conductance. We try to vary the parameters of each model so that they fit as best as possible to the data around the conductance peak, which corresponds to the 1D dispersion. The figures show that the interacting model is in good agreement with the data on both sides of the peak, with parameters $K_c = 0.704$ and $\Gamma_{\mathrm{Int}} = 0.36$ meV. In contrast, neither the non-interacting nor the constant power-law model can be made to fit the data simultaneously on both sides of the peak. In

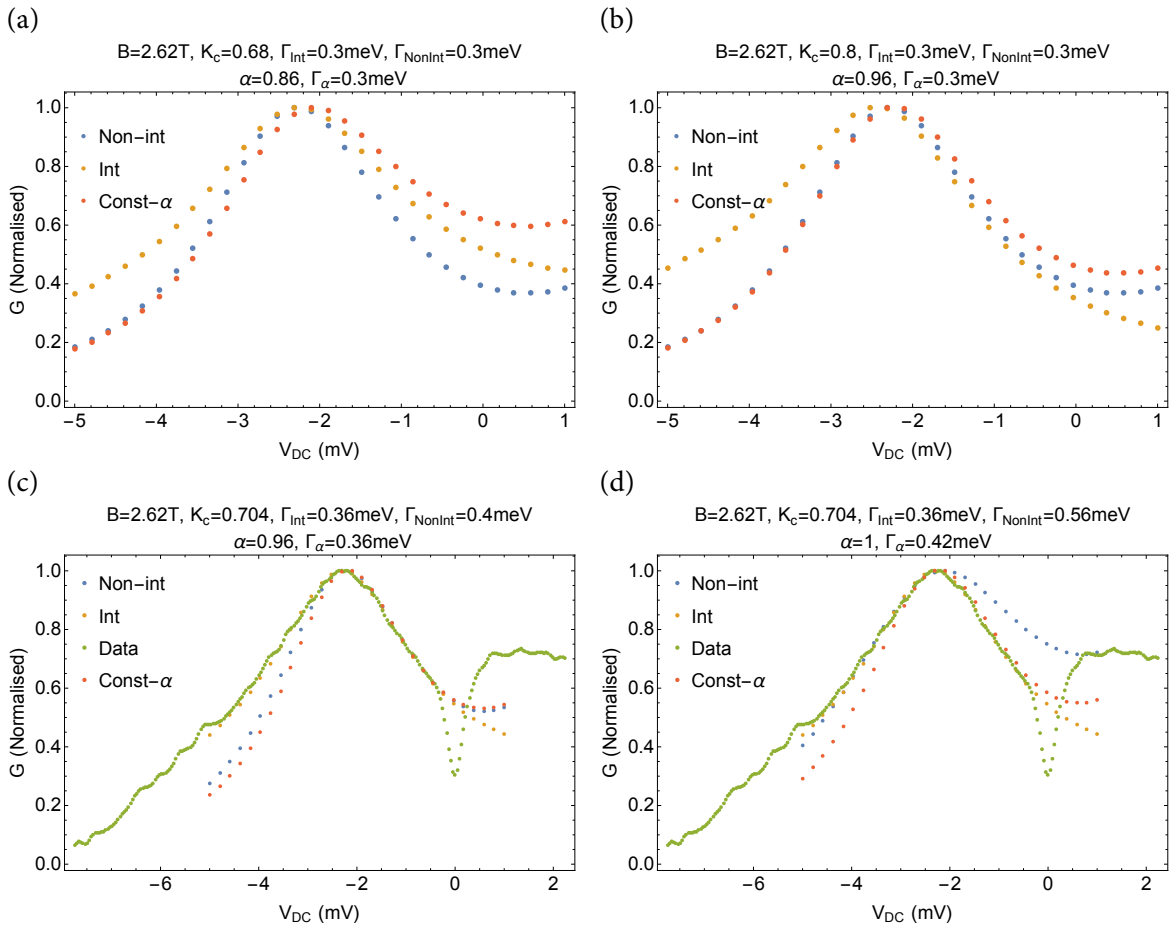


FIGURE 7.7 Comparison of tunnelling conductance calculated with the interacting, non-interacting and constant power-law model. Each model is normalised with respect to its peak conductance. The parameters used for each model is listed on top of the figures. The data series in subfigure (c) and (d) is obtained from the measurement of wafer C2617 sample on Fridge 2 at 50 mK. See main text for detailed discussions.

subfigure (c) shows that while the two models are in agreement with the data on the right hand side of the peak, they fall below the data on the left hand side. Conversely, subfigure (d) shows that while the non-interacting model is matched onto the data on the left hand side of the peak, it overshoots the data on the right hand side.

Figure 7.8 consists of three line plots which show the comparison between the cuts of data and the numerical calculations. Each plot contains four groups of curves, each comprising of three data series: the solid curve represents the cut of conductance data at a particular magnetic field (see text labels in plot). It is flanked by two dash-dotted curves, which represent the upper and lower boundaries of the uncertainties of the background elimination process, corresponding a difference of ±10% in the subtracted background conductance for Sample I and ±20% for Sample II (due to the larger background noise). Two discrete series of '•' and

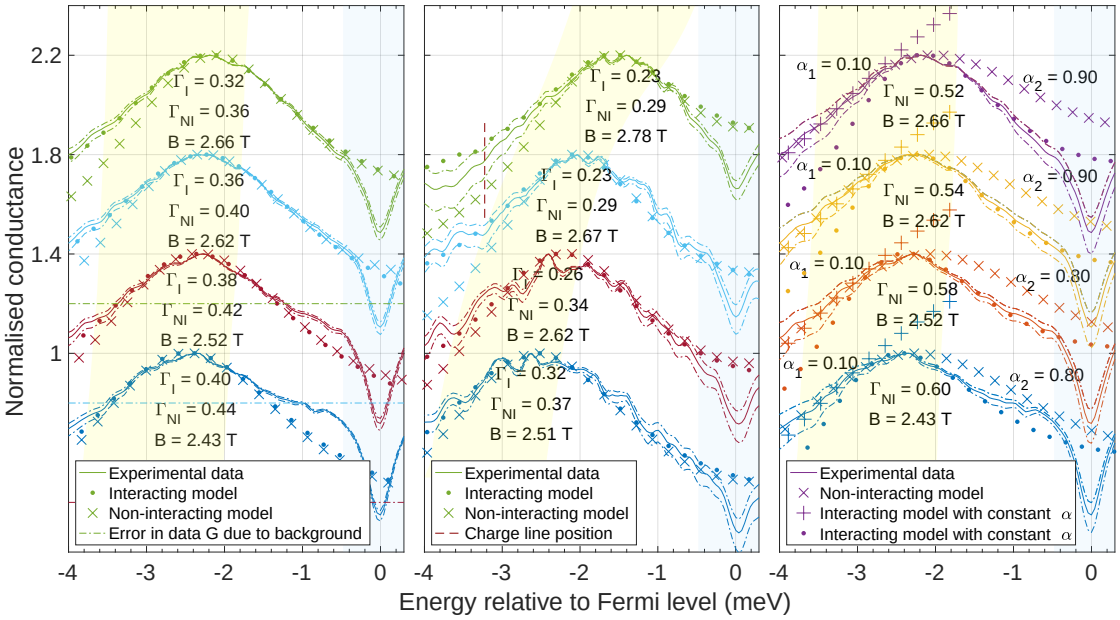


FIGURE 7.8 Comparisons between conductance data measured at fixed magnetic fields and the momentum-dependent/independent interacting and non-interacting models. The data in left and right plots are measured on Sample I and that in the middle plot is measured on Sample II. The yellow semi-transparent region represents the area where the comparison is made between the data and the model. The blue semi-transparent region represents the area where the conductance is affected by the suppression due to ZBA. See main text for the result analysis.

'×' labels are superimposed onto the data cuts, which represent the tunnelling conductance calculated by the momentum-dependent interacting model and the non-interacting model. Each data cut and model calculation is normalised to its respective maximal value in the plotted range, and their peaks, which correspond to the spectral edge, are aligned so that we can compare the line shapes. For better presentation, each group of data and calculation series is offset vertically by 0.4 normalised conductance unit from the neighbouring group. A yellow semi-transparent region can be seen in each plot, indicating the same area that is highlighted by the dotted black curve in the main conductance map.

The left and middle plots of Figure 7.8 depict conductance cuts from Sample I and II, respectively. It was found that K_c affects the skewness of the conductance peak in the model calculation, and the model which best fits the data has $K_c = 0.72 \pm 0.03$ for Sample I and $K_c = 0.69 \pm 0.03$ for Sample II. Additionally, the effective mass which best fits the data is $(0.92 \pm 0.05)m^*$ for Sample I and $(0.81 \pm 0.05)m^*$ for Sample II, where m^* is the electron effective mass. From the ratio between the spin and charge lines in Figure 7.5 and 7.6, one can extract K_c as v_s/v_c, which for Sample I comes out to be 0.76 ± 0.06 and 0.61 ± 0.06 for Sample II. The range of acceptable values for the broadening parameter, Γ_I (of the interacting

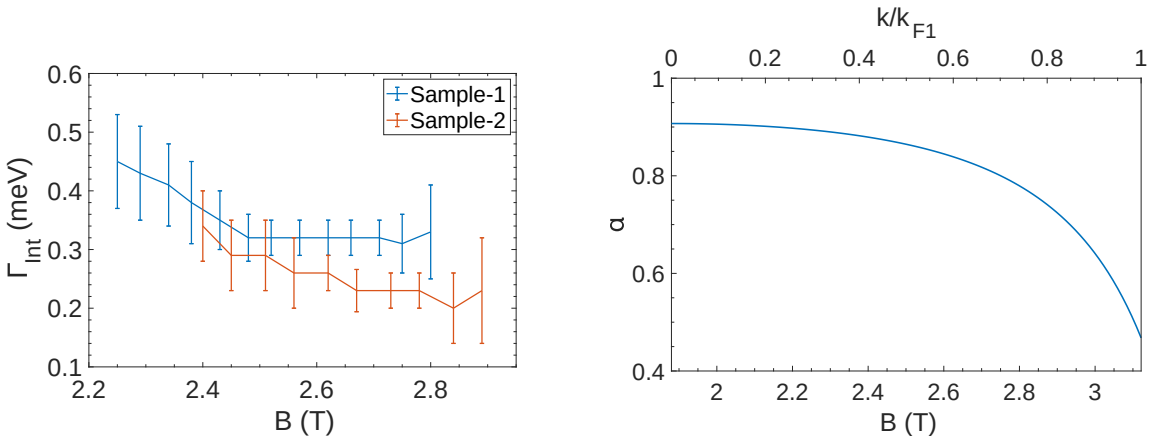


FIGURE 7.9 Left: Variation of the broadening parameter of the momentum-dependent interacting model as a function of the magnetic field. The error bars show the range of Γ values which produce the best fits to the data. Right: Variation of the power-law exponent of the interacting model, as a function of momentum and magnetic field.

model), at different $\boldsymbol{B}$ fields is shown in Figure 7.9, and the best values of both Γ_{I} and Γ_{NI} (of the non-interacting model) are listed by the text labels in Figure 7.8. Figure 7.9 shows that for both samples, Γ_{I} is minimal in a range of $\boldsymbol{B}$ values. For sample I, this occurs between 2.5 to 2.7 T, where $\Gamma_{\mathrm{I}} = 0.32 \pm 0.03$ meV; for Sample II, this occurs between 2.6 to 2.8 T, where $\Gamma_{\mathrm{I}} = 0.23 \pm 0.03$ meV. The conductance cut is taken from the $\boldsymbol{B}$ range where Γ_{I} is minimal so that the behaviour of the spectral function is not obscured by a large broadening effect. The increased broadening can be attributed to the presence of other conductance features such as the spin/charge line and the 2D parabola, which were mentioned earlier. A dashed vertical red line can be seen in the middle plot of Figure 7.8, which indicates the energy of the spin line along that particular cut. A corresponding enhancement is observed in the vicinity of the line. In both sets of results, the non-interacting model is consistently lower than the interacting model and the data to the left of the conductance peak, which is expected because there are no states below the dispersion in the non-interacting model. In contrast, the non-linear interacting model has a much slower drop-off to the left of the peak, because it allows the creation of many-body excitations in this energy range, as is shown by Figure 1.7. A good fit can be observed for both samples, between the momentum-dependent interacting model and the data. The figure also shows the non-interacting model with broadening parameter Γ_{NI} similar to Γ_{I}. Unlike the interacting model, the agreement between the non-interacting model and the data is limited to a narrow range of energies around the peak. In the third plot of Figure 7.8, comparison is made between the data from Sample I and the interacting model with constant power-law exponent. The figure shows that the exponent must be made either very small ($\alpha = 0.10$) or close to unity ($\alpha = 0.80$ or 0.90), in order to achieve a good fit to the

data. However, in either case the agreement is limited to only one side of the conductance peak. When α is small the model fits the data well to the left of the spectral edge, although it becomes divergent past the right of the data peak and does not possess a peak itself; on the other hand, when α is close to unity, the model agrees well with the data to the right of the peak, and the model possesses a peak itself. However, it significantly underestimates the conductance to the right of the data peak. The third figure also shows the non-interacting model with a Γ much larger than that of the other two plots. While the model makes good fit with the data to the left of the peak, it consistently overestimates the conductance to the right of the peak, showing that the non-interacting model cannot fully account for the observed variation of conductance. In conclusion, we find the interacting model to be consistently better than the other two models in describing the conductance around the 1D dispersion peak across a range of magnetic field. We have observed the fitted 1D-2D peaks showing variation in the spread Γ across the fitted region, and have attributed this to the presence of other dispersions in the spectral map. The lowest fitted Γ observed in our fitting, 0.32 meV for sample I and 0.23 meV for sample II, are smaller than the line width of electron-impurity scattering reported in the 2D-2D conductance study of Turner [94] ($\Gamma_{e-\mathrm{Imp}} = 0.5\,\mathrm{meV}$), and Jompol [44] ($\Gamma_{\mathrm{2D-2D}} = 0.4\,\mathrm{meV}$), suggesting that our sample wafer suffers less impurity scattering and is suited for the spectroscopy study. Instead the ability of our device to resolve dispersion features is limited by the use of the tunnel array and the quality/uniformity of the EBL.

7.4 Power law of voltage dependence of ZBA

As discussed in the last chapter, the zero bias anomaly is caused by the reduction of the linear TLL density of states at the Fermi energy. An important property of the TLL model is that all of its correlation functions behave as power laws with non-universal exponents [28]. In Figure 7.8, the suppression of the conductance caused by the ZBA is highlighted by the blue semi-transparent region around the zero DC bias. The variation of the conductance in the ZBA region can be fitted as a power-law. The conductance is modelled by the function $G = a\tilde{V}^{\beta}$. $\tilde{V}$ is the modified voltage [88] given by $\tilde{V} = \sqrt{V_{\mathrm{AC}}^2 + V_{\mathrm{DC}}^2 + \left(3k_{\mathrm{B}}T/e\right)^2}$, where V_{AC} is the AC excitation used for phase-sensitive measurement, and the last term is the equivalent voltage due to thermal energy, with T being the electron temperature.

We studied the sample conductance at five equally-spaced magnetic fields between 2 to 3.5 T, in a small DC voltage interval around the zero bias (–1 to 1 mV). At each field the conductance was measured in five repeated sweeps. Before the conductance can be fitted as a power-law, the contribution to the total conductance from outside the wire-gate array needs to be eliminated. Ideally this is done by monitoring the total tunnelling conductance

while negatively sweeping the wire-gate voltage, while holding the voltages on other gates constant, until the sample is pinched off [44]. However in the experiment run the sweep was stopped at −0.9 V to avoid damaging the e-beam gates. At this voltage the sample was not pinched off completely, so we must estimate the contribution to the total conductance from outside the wire-gate array by extrapolation. A linear fit is carried out using the portion of the conductance data highlighted in red (between $V_{\mathrm{F}} = -0.75$ to -0.9 V), and the conductance is estimated from the fitted line at the wire-gate voltage of normal device operation. We calculate the average background from several sweeps at the same $\boldsymbol{B}$ field, and subtracted this value from the conductance cuts before fitting the data. The plot at the bottom of Figure 7.12 shows such a procedure performed on a sweep taken from Sample I.

The data with the estimated background subtracted is fitted to the power-law model in Mathematica, where the exponent β and the electron temperature $\boldsymbol{T}$ are fit parameters. A constraint of $50 < \boldsymbol{T} < 120$ (in mK) was placed on the temperature parameter because we estimate the electron temperature to be within this range. As shown in Figure 7.10, the fitting is performed using a small interval of data (± 0.3 mV) around the zero bias, highlighted by the orange coloured points. The resultant parameters are displayed as the title of each fit. Another similar fit was performed, but the amplitude of the subtracted background was set to 90% of the original estimated value. The 10% reduction in the background reflects the uncertainty in our estimate of the background subtraction and the results are shown in Figure 7.11. We can see from the two figures that that the power-law and temperature parameters are very sensitive to the subtracted background. We discard the fits that have the temperature parameter equal to the boundary value. The best results are shown in the two panels in the top row of Figure 7.12. The blue region indicates the voltage range in which the fitting is carried out. From the fitting parameters we can estimate that the electron temperature was between 70 to 85 mK. The best fitting results are chosen such that the fitted temperature are consistent with the experiment and for all the conductance cuts at different $\boldsymbol{B}$ fields. The fits give an average exponent of $\beta = 0.48 \pm 0.24$, which, using the formulation of [96] which relates the exponent to the interaction parameters: $\beta = (g^{-1} - 1)/2$, where $g = K_c/K_s$ (in what is referred to as the end-tunnelling mode), implies that the Luttinger parameter is $K_c \simeq 0.50$. In comparison, Jompol [44] reported observing power-law exponent $\alpha_T = 0.48 \pm 0.03$ and $\alpha_V = 0.52 \pm 0.03$, which are based separately on temperature and voltage fittings. These are very close to the values that we have observed in our devices.

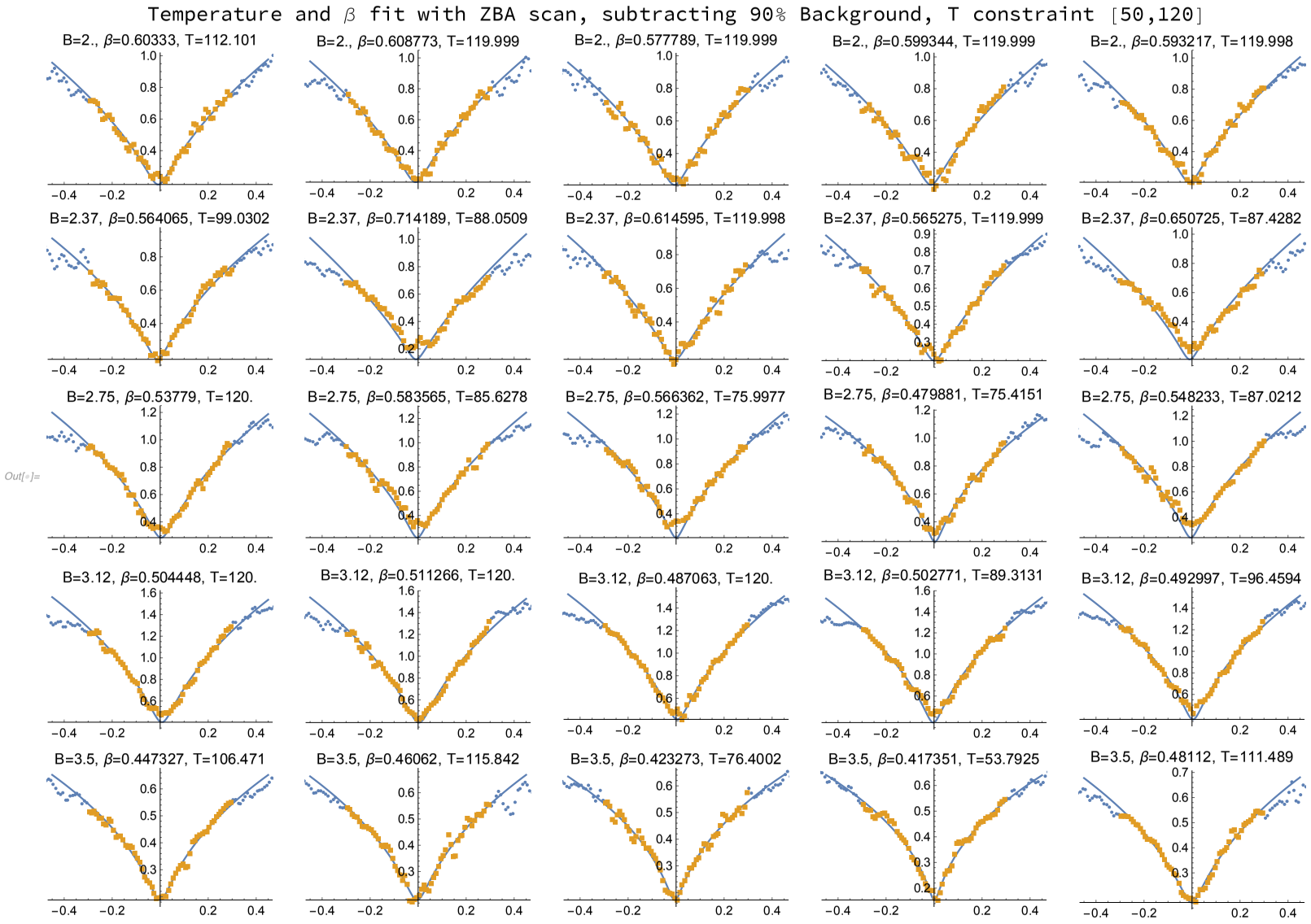


FIGURE 7.10 Results of ZBA temperature and power-law fit, performed on data measured from sample I with background conductance removed.

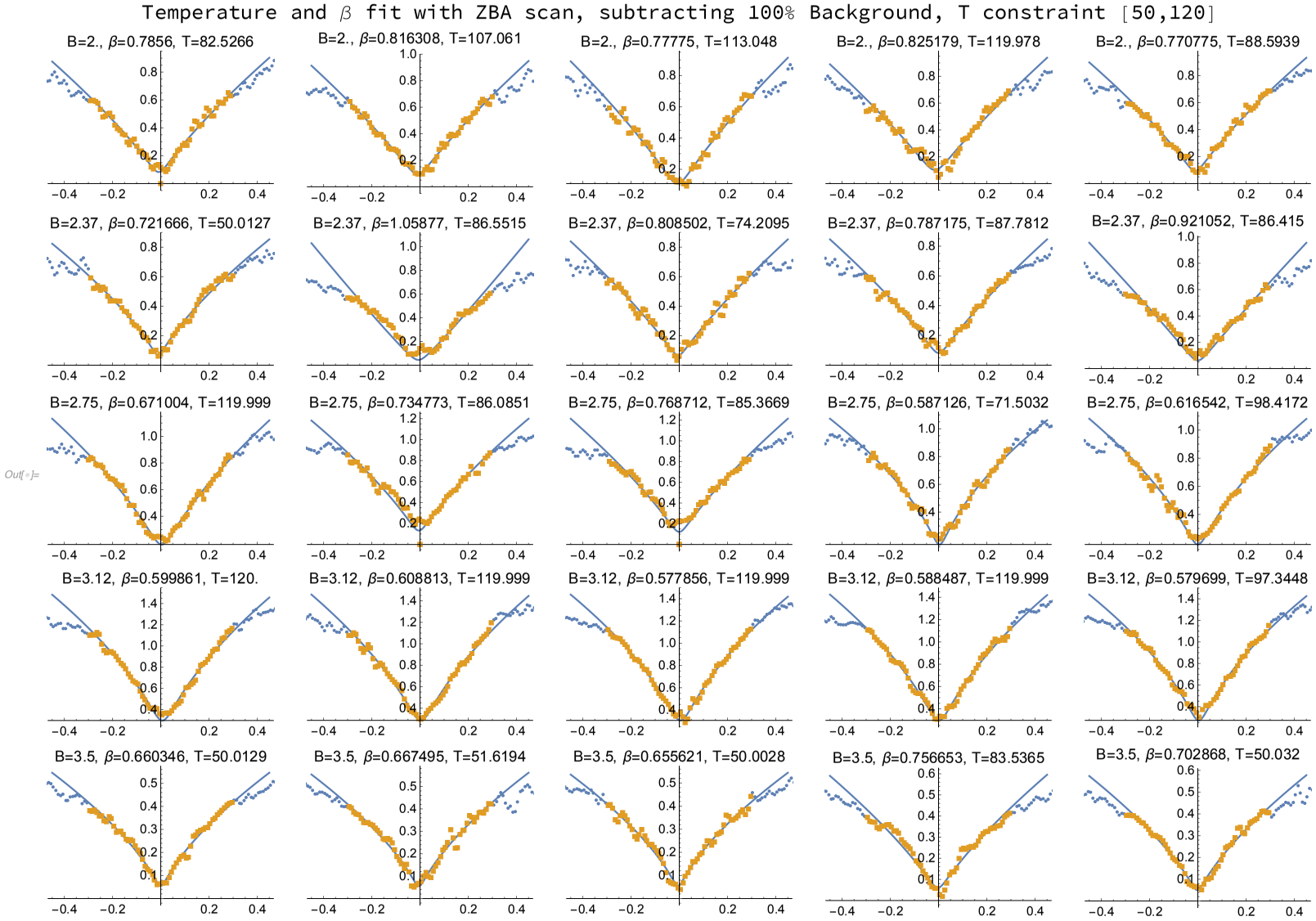


FIGURE 7.11 Results of ZBA temperature and power-law fit, performed on data measured from sample I with 90% of background conductance removed.

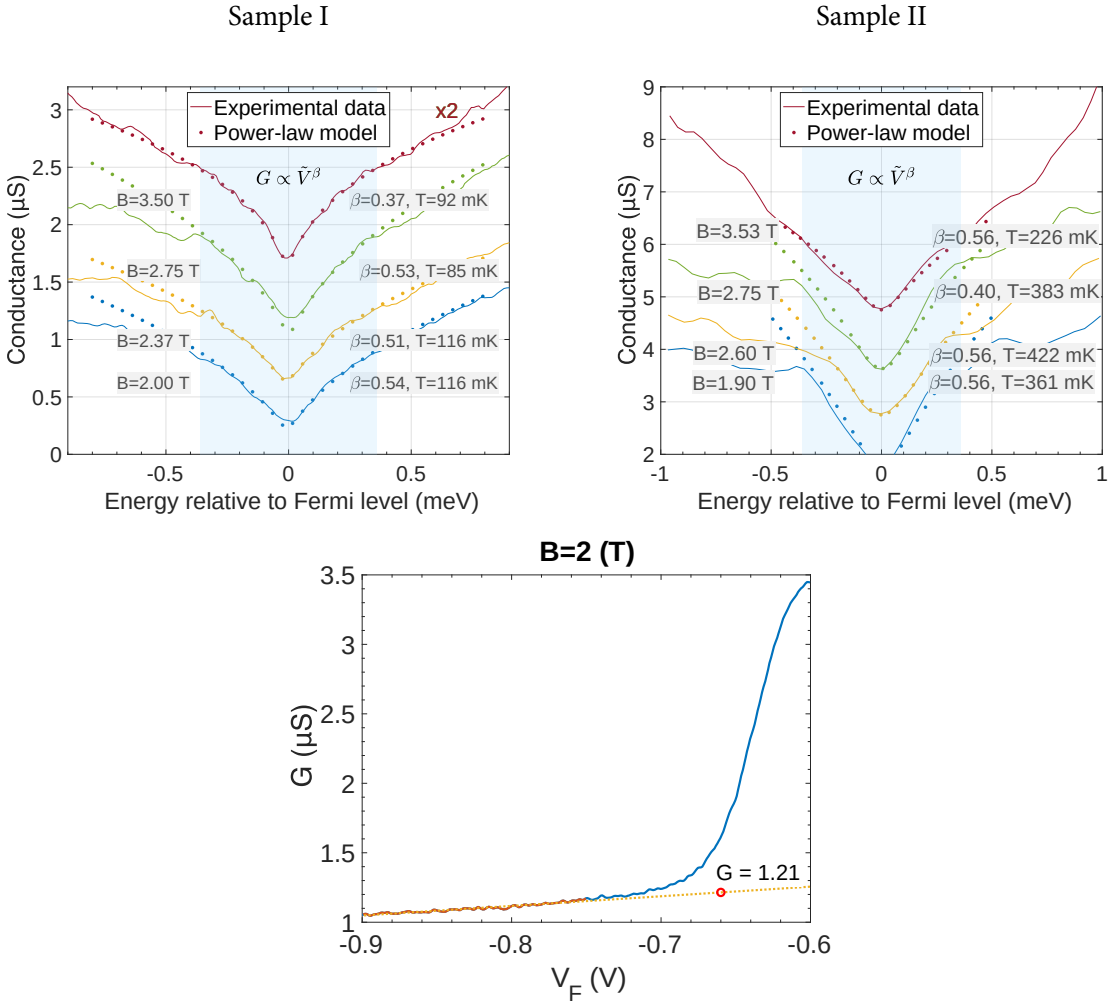


FIGURE 7.12 Left: Tunnelling conductance around the ZBA of Sample I at various $\boldsymbol{B}$ fields (solid curves), and the conductance calculated using the power-law fit (dotted curves). Right: Tunnelling data and fits around the ZBA of Sample II. Bottom: Conductance of sample when the wire gate is swept negatively. The dotted line is the linear fit of the part of conductance highlighted in red, which is used to extrapolate the tunnelling conductance from outside the wire-gate array. The extra conductance during normal device operation is read from the dotted line (in this case at $V_F = -0.66$ V, highlighted by the red circle).

Chapter 8

Conclusions and suggestions for future work

In this thesis, we started with a review of the fundamental theoretical concepts of the many-body problem, where it was shown that the description of excitations of a many-body system, and consequently, most physical observables, are closely related to a quantity known as the spectral function. We have seen that the reduced degree of freedom of a one-dimensional system results in a dynamics of interaction that is drastically different from system in two or three dimensions. The solution of the one-dimensional problem is obtained elegantly by the TLL model using the bosonisation method, which showed us that the 1D problem must be understood with a change of basis of the emergent phenomena used to describe the system. The quasiparticle picture, which so successfully extends the single-particle picture to the many-body problem in higher dimensions, must be replaced by collective modes which are separately ordered according to the charge and spin properties of the single particles.

The TLL theory suffers from a significant limitation, in that a key pre-assumption of the bosonisation method is the replacement of the parabolic single-particle dispersion by a linear function. In the first chapter of this thesis, we have discussed recent theoretical works which aimed to provide predictions outside the traditional low-energy regime. Specifically, these include the full Bethe-Ansatz diagonalisation of the interacting Hamiltonian, and the phenomenological theory of the mobile impurity model. The key results of these works have been the prediction of the existence of multiple higher-order excitation modes, and the modification of the TLL power-law singularity by a momentum-dependent exponent.

Despite the exact and distinctive predictions of the theoretical framework, it has been difficult to verify the findings experimentally. Physical observables are closely related to the spectral functions, but not as direct measures. Our experimental scheme is based on the measurement of a tunnelling conductance between a 1D and a 2D layer of electrons, whose magnitude is directly proportional to the convolution of the spectral functions of the two systems. The spectral function of the 2D electron system is relatively simple and better under-

stood than that of the 1D system, making it a good probe to study the unknown properties of a 1D system. In practice, we have built upon the successful work of the predecessor project and extended the geometric complexity of the energy-momentum-resolved tunnelling device, by developing a technique which enables the fabrication of three-dimensional EBL structures. We have used the new generation of devices to measure the data that is presented in the last two chapters of this thesis.

Analysing the gathered data, we have been able to clearly resolve, up to three 1D subbands, the parabolic dispersions of the 1D and 2D electron layers in the sample. We have demonstrated the detailed process with which the parameters of the dispersions are determined. We have been able to observe one of the key features of the linear TLL theory: the separation of spin and charge density waves. We also observed clearly the zero-bias anomaly—another feature of the linear TLL. According to theoretical prediction, the variation of the conductance as a function of the DC bias has been successfully fitted with a power law, where we have recovered similar results to the previous study. We have measured the ratio of the velocities of the two types of excitations, and found them to be similar to the observations made previously in the predecessor project. Looking beyond the low-energy limit, we consistently observed enhanced conductance beyond the high magnetic field crossing point, which we identified as the second level excitation predicted by the Bethe-Ansatz calculations. We carried out numerical calculations of the conductance, using three different forms of spectral functions for the 1D electrons, and compared the results with the measured conductance. It was found that the observed conductance agrees the most with the model of mobile impurities, which postulated the modification of the linear-TLL spectral function by a momentum-dependent power-law. We have been able to extract the Luttinger parameter K_c from the power-law fitting and the ratio of the spin and charge wave velocities, and found them to be in good agreement with each other.

Overall, this project has further demonstrated the energy-momentum-resolved tunnelling as a powerful technique for studying the interaction effects of one-dimensional electrons. However, some issues remain unsolved, which limited the better analysis of the data. The most significant issue is the appearance of fringe/interference patterns across the conductance maps which accompanies the introduction of multiple columns of wire-gate arrays. To further complicate the matter, the intensity of the patterns increases as the number of columns. This is especially troubling since reducing the wire-gate length and making a length-dependence study is of great interest, but a consequence of making shorter wire gates is increasing the number of columns. A second issue lies in the effectiveness of the parasitic region control gates. These gates were made to cover the gaps between wire-gate columns, and the effect of applying a bias on these gates has been inconsistent. While on Sample II it

had been shown to be able to induce a shift of the p-region parabola with a positive bias, on Sample I it had produced no observable effects with negative bias. Since the fringe pattern is thought to be related to conductance in the p-region [2], a systematic study and re-design of the p-region (gate) is therefore urgently required before further study can be made with the multi-column array device.

We have only carried out numerical calculations of the conductance in the single-1D-subband regime. The calculation script in fact allows the input of multiple subbands (up to four), although this functionality only expands the Hermite polynomial to the corresponding degree in the expressions of the spectral functions of the second chapter. Given the many sets of results where multiple subbands were observed, the effect of subband interaction will be an interesting theoretical problem to explore for future projects.

Finally, due to time constraints we have not been able to carry out full analysis of the temperature dependence of the interaction effects, such as the ZBA and the higher-mode excitations and the power law of the mobile-impurity model. Investigations in the theory and experimental domain on the temperature dependence remains an unexplored opportunity.

References

[1] Altland, A., Barnes, C. H. W., Hekking, F. W. J., and Schofield, A. J. (1999). Magneto-tunneling as a probe of Luttinger-liquid behavior. *Phys. Rev. Lett.*, 83:1203–1206.

[2] Anirban, A. S. (2015). *Higher-order excitations in one-dimensional electron systems*. PhD thesis, University of Cambridge.

[3] Auslaender, O. M., Steinberg, H., Yacoby, A., Tserkovnyak, Y., Halperin, B. I., Baldwin, K. W., Pfeiffer, L. N., and West, K. W. (2005). Spin-charge separation and localization in one dimension. *Science*, 308(5718):88–92.

[4] Auslaender, O. M., Yacoby, A., de Picciotto, R., Baldwin, K. W., Pfeiffer, L. N., and West, K. W. (2002). Tunneling spectroscopy of the elementary excitations in a one-dimensional wire. *Science*, 295(5556):825–828.

[5] Avinun-Kalish, M., Heiblum, M., Zarchin, O., Mahalu, D., and Umansky, V. (2005). Crossover from 'mesoscopic' to 'universal' phase for electron transmission in quantum dots. *Nature*, 436(7050):529–533.

[6] Baines, Y. D. (2010). *Kondo physics in side coupled quantum dots*. Theses, Université Joseph-Fourier - Grenoble I.

[7] Balshaw, N. and Division, O. I. G. S. R. (1996). *Practical Cryogenics: An Introduction to Laboratory Cryogenics*. Oxford Instruments, Scientific Research Division.

[8] Bardeen, J. (1961). Tunnelling from a many-particle point of view. *Phys. Rev. Lett.*, 6:57–59.

[9] Barnes, C. (2012). Quantum electronics in semiconductors.

[10] Bethe, H. (1931). Zur theorie der metalle. *Zeitschrift für Physik*, 71(3):205–226.

[11] Binney, J. and Skinner, D. (2013). *The Physics of Quantum Mechanics*. OUP Oxford.

[12] Bloch, I., Dalibard, J., and Zwerger, W. (2008). Many-body physics with ultracold gases. *Rev. Mod. Phys.*, 80:885–964.

[13] Bockrath, M., Cobden, D. H., Lu, J., Rinzler, A. G., Smalley, R. E., Balents, L., and McEuen, P. L. (1999). Luttinger-liquid behaviour in carbon nanotubes. *Nature*, 397:598 EP –.

[14] Brown, K., Nicholls, J., Tieke, B., Wiegers, S., van Bentum, P., Linfield, E., Pepper, M., Ritchie, D., and Jones, G. (1995). Tunnelling between two-dimensional electron gases up to 25 t. *Physica B: Condensed Matter*, 211(1):430 – 432. Research in High Magnetic Fields.

[15] Brown, K. M., Linfield, E. H., Ritchie, D. A., Jones, G. A. C., Grimshaw, M. P., and Pepper, M. (1994a). Resonant tunneling between parallel, two-dimensional electron gases: A new approach to device fabrication using in situ ion beam lithography and molecular beam epitaxy growth. *Applied Physics Letters*, 64(14):1827–1829.

[16] Brown, K. M., Turner, N., Nicholls, J. T., Linfield, E. H., Pepper, M., Ritchie, D. A., and Jones, G. A. C. (1994b). Tunneling between two-dimensional electron gases in a strong magnetic field. *Phys. Rev. B*, 50:15465–15468.

[17] Buckingham, M. and Fairbank, W. (1961). Chapter iii the nature of the phase-transition in liquid helium. In Gorter, C., editor, *Progress in Low Temperature Physics*, volume 3, pages 80 – 112. Elsevier.

[18] Chang, A. M. (2003). Chiral Luttinger liquids at the fractional quantum Hall edge. *Rev. Mod. Phys.*, 75:1449–1505.

[19] Demmerle, W., Smoliner, J., Berthold, G., Gornik, E., Weimann, G., and Schlapp, W. (1991). Tunneling spectroscopy in barrier-separated two-dimensional electron-gas systems. *Phys. Rev. B*, 44:3090–3104.

[20] Denlinger, J. D., Gweon, G.-H., Allen, J. W., Olson, C. G., Marcus, J., Schlenker, C., and Hsu, L.-S. (1999). Non-Fermi-liquid single particle line shape of the quasi-one-dimensional non-CDW metal $Li_{0.9}Mo_6O_{17}$: Comparison to the Luttinger liquid. *Phys. Rev. Lett.*, 82:2540–2543.

[21] Dupuis, N. (2014). Fermi liquid theory. *Laboratoire de Physique Theorique de la Matiere Condensee Website, Sorbonne Université*.

[22] Egger, R. and Gogolin, A. O. (1997). Effective low-energy theory for correlated carbon nanotubes. *Phys. Rev. Lett.*, 79:5082–5085.

[23] Eisenstein, J. P., Gramila, T. J., Pfeiffer, L. N., and West, K. W. (1991a). Probing a two-dimensional fermi surface by tunneling. *Phys. Rev. B*, 44:6511–6514.

[24] Eisenstein, J. P., Pfeiffer, L. N., and West, K. W. (1991b). Field-induced resonant tunneling between parallel two-dimensional electron systems. *Applied Physics Letters*, 58(14):1497–1499.

[25] Eisenstein, J. P., Pfeiffer, L. N., and West, K. W. (1992). Coulomb barrier to tunneling between parallel two-dimensional electron systems. *Phys. Rev. Lett.*, 69:3804–3807.

[26] Ford, C. J. B., Thornton, T. J., Newbury, R., Pepper, M., Ahmed, H., Peacock, D. C., Ritchie, D. A., Frost, J. E. F., and Jones, G. A. C. (1989). Electrostatically defined heterojunction rings and the Aharonov-Bohm effect. *Appl. Phys. Lett.*, 54:21 – 23.

[27] Giamarchi, T. (2004). *Quantum Physics in One Dimension*. International Series of Monographs on Physics. Clarendon Press.

[28] GIAMARCHI, T. (2012). Some experimental tests of Tomonaga–Luttinger liquids. *International Journal of Modern Physics B*, 26(22):1244004.

[29] Giamarchi, T. (2016). One-dimensional physics in the 21st century. *Comptes Rendus Physique*, 17(3):322 – 331. Physique de la matière condensée au XXIe siècle: l'héritage de Jacques Friedel.

[30] Girgis, E., Liu, J., and Benkhedar, M. L. (2006). Fabrication of metallic air bridges using multiple-dose electron beam lithography. *Applied Physics Letters*, 88(20).

[31] Glazman, L. I., Ruzin, I. M., and Shklovskii, B. I. (1992). Quantum transport and pinning of a one-dimensional wigner crystal. *Phys. Rev. B*, 45:8454–8463.

[32] Grigera, S. A., Schofield, A. J., Rabello, S., and Si, Q. (2004). Momentum-resolved tunneling between a Luttinger liquid and a two-dimensional electron gas. *Phys. Rev. B*, 69:245109.

[33] Grioni, M., Pons, S., and Frantzeskakis, E. (2008). Recent ARPES experiments on quasi-1D bulk materials and artificial structures. *Journal of Physics: Condensed Matter*, 21(2):023201.

[34] Gurman, I., Sabo, R., Heiblum, M., Umansky, V., and Mahalu, D. (2016). Dephasing of an electronic two-path interferometer. *Phys. Rev. B*, 93:121412.

[35] Hager, J., Matzdorf, R., He, J., Jin, R., Mandrus, D., Cazalilla, M. A., and Plummer, E. W. (2005). Non-fermi-liquid behavior in quasi-one-dimensional $Li_{0.9}Mo_6O_{17}$. *Phys. Rev. Lett.*, 95:186402.

[36] Haldane, F. D. M. (1981). 'Luttinger liquid theory' of one-dimensional quantum fluids. i. properties of the Luttinger model and their extension to the general 1d interacting spinless fermi gas. *Journal of Physics C: Solid State Physics*, 14(19):2585–2609.

[37] Halperin, B. I. (1982). Quantized hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential. *Phys. Rev. B*, 25:2185–2190.

[38] Harrison, S. (2016). *Exploring and Exploiting Charge-Carrier Confinement in Semiconductor Nanostructures*. PhD thesis, Lancaster University.

[39] Imambekov, A. and Glazman, L. I. (2009a). Phenomenology of one-dimensional quantum liquids beyond the low-energy limit. *Phys. Rev. Lett.*, 102:126405.

[40] Imambekov, A. and Glazman, L. I. (2009b). Universal theory of nonlinear Luttinger liquids. *Science*, 323(5911):228–231.

[41] Jeffrey, A., Littlewood, A., and Sergeevič, V. V. (1971). *Equations of mathematical physics*. Dekker.

[42] Jérome, D. (2004). Organic conductors: From charge density wave ttf-tcnq to superconducting (tmtsf)2pf6. *Chemical Reviews*, 104(11):5565–5592.

[43] Jin, Y., Tsyplyatyev, O., Moreno, M., Anthore, A., Tan, W. K., Griffiths, J. P., Farrer, I., Ritchie, D. A., Glazman, L. I., Schofield, A. J., and Ford, C. J. B. (2019). Momentum-dependent power law measured in an interacting quantum wire beyond the Luttinger limit. *Nature Communications*, 10(1):2821.

[44] Jompol, Y. (2008). *Probing Tomonaga-Luttinger Liquid behaviour and spin-charge separation in GaAs quantum wires*. Phd theses, Department of Physics, University of Cambridge.

[45] Jompol, Y., Ford, C. J. B., Griffiths, J. P., Farrer, I., Jones, G. A. C., Anderson, D., Ritchie, D. A., Silk, T. W., and Schofield, A. J. (2009). Probing spin-charge separation in a Tomonaga-Luttinger liquid. *Science*, 325(5940):597–601.

[46] Kane, C., Balents, L., and Fisher, M. P. A. (1997). Coulomb interactions and mesoscopic effects in carbon nanotubes. *Phys. Rev. Lett.*, 79:5086–5089.

[47] Kane, C. L. and Fisher, M. P. A. (1992a). Transmission through barriers and resonant tunnelling in an interacting one-dimensional electron gas. *Phys. Rev. B*, 46:15233–15262.

[48] Kane, C. L. and Fisher, M. P. A. (1992b). Transport in a one-channel Luttinger liquid. *Phys. Rev. Lett.*, 68:1220–1223.

[49] Kane, M. J., Apsley, N., Anderson, D. A., Taylor, L. L., and Kerr, T. (1985). Parallel conduction in GaAs/Al_xGa_{1-x} modulation doped heterojunctions. *Journal of Physics C: Solid State Physics*, 18(29):5629.

[50] Kardynał, B., Barnes, C. H. W., Linfield, E. H., Ritchie, D. A., Brown, K. M., Jones, G. A. C., and Pepper, M. (1996). Direct measurement of the band structure of a one-dimensional surface superlattice. *Phys. Rev. Lett.*, 76:3802–3805.

[51] Kardynał, B., Barnes, C. H. W., Linfield, E. H., Ritchie, D. A., Nicholls, J. T., Brown, K. M., Jones, G. A. C., and Pepper, M. (1997). Magnetotunneling spectroscopy of one-dimensional wires. *Phys. Rev. B*, 55:R1966–R1969.

[52] Kardynal/, B., Linfield, E. H., Ritchie, D. A., Brown, K. M., Barnes, C. H. W., Jones, G. A. C., and Pepper, M. (1996). Equilibrium tunneling between two-dimensional and quasi-one-dimensional electron gases in devices fabricated by in situ focused ion beam lithography. *Applied Physics Letters*, 68(6):826–828.

[53] Khalid, A., Li, C., Grant, J., Saha, S., Ferguson, S., and Cumming, D. R. (2012). Simple e-beam air-bridge technology for mm-wave applications. *Microelectronic Engineering*, 98:262 – 265. Special issue {MNE} 2011 - Part {II}.

[54] Khodas, M., Pustilnik, M., Kamenev, A., and Glazman, L. I. (2007). Fermi-Luttinger liquid: Spectral function of interacting one-dimensional fermions. *Phys. Rev. B*, 76:155402.

[55] Korepin, V. E., Bogoliubov, N. M., and Izergin, A. G. (1993). *Quantum Inverse Scattering Method and Correlation Functions*. Cambridge Monographs on Mathematical Physics. Cambridge University Press.

[56] Kuznetsov, A. and Kwaśnicki, M. (2017). Minimal Hermite-type eigenbasis of the discrete Fourier transform. *arXiv e-prints*, page arXiv:1706.08740.

[57] Lebed, A. G. (2008). *The physics of organic superconductors and conductors*. Springer.

[58] Li, M., Chen, L., and Chou, S. Y. (2001). Direct three-dimensional patterning using nanoimprint lithography. *Applied Physics Letters*, 78(21):3322–3324.

[59] Lounis, S. (2014). Theory of scanning tunneling microscopy.

[60] Macks, L., Barnes, C., Nicholls, J., Tribe, W., Ritchie, D., Rose, P., Linfield, E., and Pepper, M. (2000). A direct measurement of the effects of fermi energy oscillations in quasi-1d systems. *Physica E: Low-dimensional Systems and Nanostructures*, 6(1):518 – 521.

[61] Mahan, G. D. (2014). *Many-particle physics*. Springer.

[62] Matveev, K. A. and Glazman, L. I. (1993). Coulomb blockade of tunneling into a quasi-one-dimensional wire. *Phys. Rev. Lett.*, 70:990–993.

[63] Meden, V. and Schönhammer, K. (1992). Spectral functions for the Tomonaga-Luttinger model. *Phys. Rev. B*, 46:15753–15760.

[64] Moreno, M., Ford, C. J. B., Jin, Y., Griffiths, J. P., Farrer, I., Jones, G. A. C., Ritchie, D. A., Tsyplyatyev, O., and Schofield, A. J. (2016). Nonlinear spectra of spinons and holons in short GaAs quantum wires. *Nature Communications*, 7(1):12784.

[65] Negele, J. and Orland, H. (1988). *Quantum many-particle systems*. Frontiers in physics. Addison-Wesley Pub. Co.

[66] Nield, S. A., Nicholls, J. T., Tribe, W. R., Simmons, M. Y., and Ritchie, D. A. (2000). Independent contacts to a double-electron gas using mesoscopic surface gates. *Journal of Applied Physics*, 87(8):4036–4038.

[67] Nikolajsen, J. R. (2013). Edge states and contacts in the quantum hall effect. Master's thesis, FACULTY OF SCIENCE UNIVERSITY OF COPENHAGEN.

[68] NOZIÈRES, P. and DE DOMINICIS, C. T. (1969). Singularities in the X-ray absorption and emission of metals. iii. one-body theory exact solution. *Phys. Rev.*, 178:1097–1107.

[69] NOZIÈRES, P., GAVORET, J., and ROULET, B. (1969). Singularities in the X-ray absorption and emission of metals. ii. self-consistent treatment of divergences. *Phys. Rev.*, 178:1084–1096.

[70] Odashima, M., G. Prado, B., and Vernek, E. (2016). Pedagogical introduction to equilibrium Green's functions: Condensed-matter examples with numerical implementations. *Revista Brasileira de Ensino de Física*, 39:2017.

[71] Pfeiffer, L., Yacoby, A., Stormer, H., Baldwin, K., Hasen, J., Pinczuk, A., Wegscheider, W., and West, K. (1997). Transport and optics in quantum wires fabricated by mbe overgrowth on the (110) cleaved edge. *Microelectronics Journal*, 28(8):817 – 823. Novel Index Semiconductor Surfaces: Growth, Characterization and Devices.

[72] Pobell, F. (2007). *Matter and methods at low temperatures*. Springer.

[73] Pustilnik, M., Khodas, M., Kamenev, A., and Glazman, L. I. (2006). Dynamic response of one-dimensional interacting fermions. *Phys. Rev. Lett.*, 96:196405.

[74] Rooks, M. J. (2002). Low stress development of poly(methylmethacrylate) for high aspect ratio structures. *Journal of Vacuum Science Technology B: Microelectronics and Nanometer Structures*, 20:2937.

[75] ROULET, B., GAVORET, J., and NOZIÈRES, P. (1969). Singularities in the X-ray absorption and emission of metals. i. first-order parquet calculation. *Phys. Rev.*, 178:1072–1083.

[76] Schmidt, T. L., Imambekov, A., and Glazman, L. I. (2010a). Fate of 1d spin-charge separation away from fermi points. *Phys. Rev. Lett.*, 104:116403.

[77] Schmidt, T. L., Imambekov, A., and Glazman, L. I. (2010b). Spin-charge separation in one-dimensional fermion systems beyond Luttinger liquid theory. *Phys. Rev. B*, 82:245104.

[78] Schofield, A. J. (1999). Non-fermi liquids. *Contemporary Physics*, 40(2):95–115.

[79] Schrieffer, J. R., Scalapino, D. J., and Wilkins, J. W. (1963). Effective tunneling density of states in superconductors. *Phys. Rev. Lett.*, 10:336–339.

[80] Schwartz, A., Dressel, M., Grüner, G., Vescoli, V., Degiorgi, L., and Giamarchi, T. (1998). On-chain electrodynamics of metallic $(TMTSF)_2x$ salts: Observation of Tomonaga-Luttinger liquid response. *Phys. Rev. B*, 58:1261–1271.

[81] Shytov, A. V., Glazman, L. I., and Starykh, O. A. (2003). Zeeman splitting of zero-bias anomaly in Luttinger liquids. *Phys. Rev. Lett.*, 91:046801.

[82] Simons, B. (2019). *Lecture Notes on Quantum Condensed Matter Field Theory*, volume 2. TCM Group Website, Cavendish Laboratory, University of Cambridge.

[83] Smith, S. (2007). An interesting fourier transform - 1/f noise.

[84] Smoliner, J., Demmerle, W., Berthold, G., Gornik, E., Weimann, G., and Schlapp, W. (1989). Momentum conservation in tunneling processes between barrier-separated 2d-electron-gas systems. *Phys. Rev. Lett.*, 63:2116–2119.

[85] Smoliner, J., Gornik, E., and Weimann, G. (1988). Depletion charge measurements by tunneling spectroscopy gaas-gaalas field-effect transistors. *Applied Physics Letters*, 52(25):2136–2138.

[86] Sze, S. M. and Ng, K. K. (2007). *Physics of semiconductor devices*. John Wiley & Sons.

[87] Sólyom, J. (1979). The fermi gas model of one-dimensional conductors. *Advances in Physics*, 28(2):201–303.

[88] Tarucha, S., Honda, T., and Saku, T. (1995). Reduction of quantized conductance at low temperatures observed in 2 to 10 μm-long quantum wires. *Solid State Communications*, 94(6):413 – 418.

[89] Tong, D. (2016). Lectures on the quantum hall effect.

[90] Tsyplyatyev, O. (2018). *Interacting quantum systems in one dimension at arbitrary energy*. Habilitationsschrift. Goethe Universität Frankfurt.

[91] Tsyplyatyev, O. and Schofield, A. J. (2014). Spectral-edge mode in interacting one-dimensional systems. *Phys. Rev. B*, 90:014309.

[92] Tsyplyatyev, O., Schofield, A. J., Jin, Y., Moreno, M., Tan, W. K., Anirban, A. S., Ford, C. J. B., Griffiths, J. P., Farrer, I., Jones, G. A. C., and Ritchie, D. A. (2016). Nature of the many-body excitations in a quantum wire: Theory and experiment. *Phys. Rev. B*, 93:075147.

[93] Tsyplyatyev, O., Schofield, A. J., Jin, Y., Moreno, M., Tan, W. K., Ford, C. J. B., Griffiths, J. P., Farrer, I., Jones, G. A. C., and Ritchie, D. A. (2015). Hierarchy of modes in an interacting one-dimensional system. *Phys. Rev. Lett.*, 114:196401.

[94] Turner, N., Nicholls, J. T., Linfield, E. H., Brown, K. M., Jones, G. A. C., and Ritchie, D. A. (1996). Tunneling between parallel two-dimensional electron gases. *Phys. Rev. B*, 54:10614–10624.

[95] Voit, J. (1993a). Charge-spin separation and the spectral properties of Luttinger liquids. *Phys. Rev. B*, 47:6740–6743.

[96] Voit, J. (1993b). Charge-spin separation and the spectral properties of Luttinger liquids. *Phys. Rev. B*, 47:6740–6743.

[97] Voit, J. (1995). One-dimensional fermi liquids. *Reports on Progress in Physics*, 58(9):977–1116.

[98] Voit, J. (2000). A brief introduction to Luttinger liquids. *AIP Conference Proceedings*, 544(1):309–318.

[99] Wang, F., Mo, S.-K., Allen, J. W., Kim, H.-D., He, J., Jin, R., Mandrus, D., Sekiyama, A., Tsunekawa, M., and Suga, S. (2006). Case for bulk nature of spectroscopic Luttinger liquid signatures observed in angle-resolved photoemission spectra of $Li_{0.9}Mo_6O_{17}$. *Phys. Rev. B*, 74:113107.

[100] Yacoby, A., Heiblum, M., Umansky, V., Shtrikman, H., and Mahalu, D. (1994). Unexpected periodicity in an electronic double slit interference experiment. *Phys. Rev. Lett.*, 73:3149–3152.

[101] Yasin, S., Hasko, D., and Ahmed, H. (2002). Comparison of MIBK/IPA and water/IPA as PMMA developers for electron beam nanolithography. *Microelectronic Engineering*, 6162:745 – 753. Micro- and Nano-Engineering 2001.

[102] Yasin, S., Hasko, D. G., and Ahmed, H. (2001). Fabrication of <5 nm width lines in poly(methylmethacrylate) resist using a water:isopropyl alcohol developer and ultrasonically-assisted development. *Applied Physics Letters*, 78(18):2760–2762.